

# **Systematic understanding of the DNA repair mechanisms induced by Cas9 mediated cleavage.**

By

**Rhythm Phutela**

(AcSIR Registration Number **10BB18A02011**)

A thesis submitted to the Academy of Scientific & Innovative Research for the award of the degree of

DOCTOR OF PHILOSOPHY  
In  
SCIENCE

Under the supervision of Dr. Debojyoti Chakraborty



**CSIR-INSTITUTE OF GENOMICS AND INTEGRATIVE BIOLOGY**

**Council of Scientific & Industrial Research South Campus, CRRI  
Campus, Mathura Road, Delhi, 110020**



**Academy of Scientific and Innovative Research**

AcSIR Headquarters, CSIR-HRDC campus Sector 19, Kamla Nehru Nagar  
Ghaziabad, U.P. 201002, India

**July 2024**

## Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled, "**Systematic understanding of the DNA repair mechanisms induced by Cas9 mediated cleavage**" submitted by Rhythm Phutela to the Academy of Scientific and Innovative Research (AcSIR), in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy in Science, embodies original research work carried-out by the student. We further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material(s) obtained from other source(s) and used in this research work has/have been duly acknowledged in the thesis. Image(s), illustration(s), figure(s), table(s) etc., used in the thesis from other source(s), have also been duly cited and acknowledged.

(Signature of Student)  
Rhythm Phutela  
14/07/2024

(Signature of Supervisor)  
Dr. Debojyoti Chakraborty  
14/07/2024

## STATEMENTS OF ACADEMIC INTEGRITY

I, Rhythm Phutela, a Ph.D. student of the Academy of Scientific and Innovative Research (AcSIR) with Registration No. 10BB18A02011 hereby undertake that, the thesis entitled “Molecular dissection of DNA targeting properties of FnCas9 for monogenic disease detection and correction” has been prepared by me and that the document reports original work carried out by me and is free of any plagiarism in compliance with the UGC Regulations on “Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)” and the CSIR Guidelines For “Ethics in Research and in Governance (2020)”.



(Signature of Student)

Rhythm Phutela

New Delhi

14/07/2024

It is hereby certified that the work done by the student, under my/our supervision, is plagiarism-free in accordance with the UGC Regulations on “*Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions (2018)*” and the CSIR Guidelines For “*Ethics in Research and in Governance (2020)*”.

(Signature of Supervisor)

Dr. Debojyoti Chakraborty

New Delhi

14/07/2024

# Index

Acknowledgments .....	I
Vernacular abstract .....	II
Abbreviations .....	III
List of Figures .....	IV
List of Tables .....	V
Chapter 1: Overview and principle of CRISPR-Cas system.....	6
1.1 Overview of CRISPR Cas system .....	7
1.2 Classification of CRISPR system .....	9
1.3 Repair pathways induced by CRISPR Cas DNA cleavage .....	10
1.4 Factors affecting DNA repair outcomes induced by CRISPR Cas9 cleavage .....	13
1.5 Regulation of the trade-off between NHEJ and HDR .....	14
1.6 Aims and objectives .....	23
1.7 Bibliography .....	24
Chapter 2: To investigate the structure of DNA breaks generated by FnCas9/en1 FnCas9 upon DNA cleavage .....	32
2.1 Background.....	33
2.2 Methods and Materials.....	34
2.2.1 Plasmid Construction .....	34
2.2.2 Protein purification .....	34
2.2.3 <i>In vitro</i> transcription .....	35
2.2.4 <i>In vitro</i> Cleavage (IVC) assay.....	35
2.2.5 Cloning of IVC products .....	35
2.2.6 Purification of the Cas9 cut linear products .....	35
2.2.7 Uptake of exogenous templates by bacterial and mammalian cells .....	36
2.2.8 Guide RNA cloning.....	36
2.2.9 Sanger Sequencing.....	36
2.2.10 Data Analysis .....	37
2.2.11 Cell culture and electroporation.....	37
2.2.12 Amplicon Sequencing.....	37
2.2.13 Statistical Analysis .....	39
2.3 Results.....	39
2.3.1 en1FnCas9 generates sticky DNA ends during <i>in-vitro</i> and in-cellulo conditions .....	39
2.3.2 Staggered DNA ends are processed differently than blunt ends in bacterial and mammalian cells .....	43

2.4 Discussion.....	48
2.5 Bibliography .....	49
Chapter 3: To comprehend the repair proteins recruitment triggered upon en1FnCas9 and SpCas9 DNA cleavage .....	52
3.1 Background.....	53
3.2 Methods and Materials.....	54
3.2.1 Plasmid Construction .....	54
3.2.2 Protein purification .....	54
3.2.3 <i>In-vitro</i> transcription.....	55
3.2.4 <i>In-vitro</i> Cleavage (IVC) assay.....	55
3.2.5 Purification of the Cas9 cut linear products .....	55
3.2.6 Electroporation of AX227 pooled library in bacterial cells .....	55
3.2.7 gRNA representation analysis in AX227 CRISPRi library .....	56
3.2.8 Cell culture and transfection .....	56
3.2.9 Construction of the dspCas9 KRAB stable cell lines.....	56
3.2.10 qRT-PCR .....	57
3.2.11 Western blot.....	57
3.2.12 Immunofluorescence (IF) .....	58
3.2.13 CRISPRi dspCas9 KRAB HEK293T cells .....	58
3.2.14 Uptake of exogenous templates by CRISPRi mammalian cells .....	59
3.2.15 Repair-seq-sequencing library preparation .....	59
3.3 Results.....	61
3.3.1 Stably expressing dSpCas9 KRAB CRISPRi HEK293T cells .....	61
3.3.2 Investigating differential blunt and staggered DNA repair factors via single cell RNA sequencing .....	65
3.4 Discussion.....	75
3.5 Bibliography .....	77
Chapter 4: To examine the repair outcomes mediated by en1 FnCas9 and SpCas9 at different loci....	79
4.1 Background.....	80
4.2 Methods and Materials.....	81
4.2.1 Plasmid Construction .....	81
4.2.2 Protein purification .....	81
4.2.3 <i>In- vitro</i> transcription.....	82
4.2.4 <i>In-vitro</i> Cleavage (IVC) assay.....	82
4.2.5 Construction of the Cas9 stable cell lines .....	83
4.2.6 Real Time PCR .....	83

4.2.7 Western blot.....	83
4.2.8 Viability and apoptotic assays for Cas9 stable cell lines .....	84
4.2.9 γ-H2AX staining .....	85
4.2.10 Electroporation in HEK293T cells .....	85
4.2.11 Culuring of hESCs.....	86
4.2.12 Transfection of Cas9 editing components using lipofectamine .....	86
4.2.13 Immunostaining of Human stem cells.....	87
4.2.14 Cell culture of mESCs .....	87
4.2.15 CRISPR/Cas9 editing in mouse zygotes .....	88
4.2.16 Amplicon sequencing .....	89
4.2.17 Long range nanopore sequencing and analysis of large deletions .....	90
4.2.18 Digenome sequencing in mESCs .....	91
4.3 Results.....	92
4.3.1 Stably expressing en1FnCas9 mediated staggered cleavage leads to higher HDR knockin rates at LMNA and FASN locus .....	92
4.3.2 Staggered cutting en1 FnCas9 suppresses NHEJ and makes homologous pathways dominant	97
4.3.3 en1 FnCas9 mediated HDR editing in human stem cells.....	102
4.3.4 en1 FnCas9 mediated editing in mouse embryos.....	105
4.3.5 Genome wide Off- target binding analysis by en1 FnCas9 .....	107
4.4 Discussion.....	109
4.5 Bibliography .....	110
Chapter 5: Summary and Future Perspectives.....	114
5.1 Concluding remark .....	115
5.2 Avenues to explore .....	116
5.3 Bibliography .....	118
Appendix.....	120
CSIR-800 Societal Project PAN India Sero- Epidemiological Survey (Phenome-India Cohort) for SARS-CoV2.....	129
Abstract.....	138
Publications .....	139

*Dedicated to my parents  
Especially my MOM*

## Acknowledgments

*Failure after long perseverance is much grander than never having a striving good enough to be called a failure*

George Eliot

This chapter stands as a heartfelt tribute to the unwavering support, guidance, and camaraderie I have been fortunate to experience throughout my PhD journey. The pursuit of this degree has been nothing short of an extraordinary rollercoaster ride, filled with invaluable lessons, memorable moments, and lasting connections. Within the pages of this thesis, I wish to express my deepest gratitude to the many people—mentors, colleagues, friends, and family—who have walked alongside me during this once-in-a-lifetime journey.

First and foremost, I am profoundly grateful to my supervisor, **Dr. Debojyoti Chakraborty**, for his constant support, insightful mentorship, and unwavering encouragement, all of which have been pivotal to my growth as a researcher. It has been a true privilege to learn under his guidance. I deeply admire not only his scientific brilliance but also his remarkable qualities as a person. I am equally indebted to **Dr. Souvik Maiti**, the Director of CSIR-IGIB, whose approachability and support—both within and beyond science—have been invaluable, despite his demanding schedule.

I would also like to sincerely thank the members of my Doctoral Advisory Committee—**Dr. Lipi Thukral, Dr. Manikandan, and Dr. T.N. Vivek**—for their constructive feedback and critical insights. My gratitude extends to **Dr. Soumya Sinha Roy, Dr. Kaushik Chakraborty, Dr. Mary Ekka**, and my extended lab family for fostering such a collaborative and supportive environment. A special mention goes to Mukesh Bhaiya, whose tireless efforts ensured the seamless functioning of our lab.

Collaboration has played a crucial role in shaping my thesis, and I was incredibly fortunate to work with **Dr. Kathy Niakan at the University of Cambridge**, thanks to the EMBO Scientific Exchange Program in the fall of 2023. I am deeply grateful to Dr. Niakan for the opportunity to work in her lab with both freedom and trust. I will always cherish our engaging conversations,

including the behind-the-scenes stories of the iconic illustration that accompanied her groundbreaking work as the first scientist to receive regulatory approval to edit human embryo genomes. I am lucky to have found not only a mentor but also a friend in her. My time in the Niakan lab was especially enriching thanks to the incredible support from **Oliver Bower**, **Dr. Desislava Staneva**, **Dr. Afshan McCarthy**, **Dr. Riley McMahon**, **Dr. Katarina Harasimov**, and **Q. Huang**, who made me feel right at home. The warmth of my welcome party in Cambridge and my farewell gatherings are memories I will always treasure.

I am also thankful to **Dr. Rajesh Pandey** for the opportunity to carry out nanopore and single-cell RNA sequencing under his guidance. Special thanks to Kriti and Jyoti for their support with RNA sequencing experiments, Dr. Partha for single-cell RNA analysis, and Dr. Ranjit for nanopore data analysis.

One of the most transformative moments of my PhD came with fieldwork—realizing that impactful science happens beyond the lab. I am grateful to have joined Dr. Debo and Dr. Souvik on visits to schools in Raipur to evaluate our CRISPR-Cas9-based diagnostic kit for Sickle Cell Anaemia. It has been an honor to contribute to developing FELUDA, a cost-effective, point-of-care diagnostic assay for both Sickle Cell Anaemia and COVID-19.

**The heartbeat of my PhD has been the friendships I've built with my labmates. I am deeply thankful to my wonderful seniors—Dr. Azhar, Dr. Amrita, Dr. Arpita, Dr. Asgar, Dr. Deepanjan, Dr. Kanan, Dr. Kanika, Dr. Manish, Dr. Monika, Dr. Jaya, Dr. Isha, and Namrata—for their constant guidance. Special love and gratitude to Dr. Riya, Dr. Meghali, Dr. Sundaram, and Dr. Saumya, who were like mentors in both science and life. My thanks extend to all past and present lab members, who have been part of this incredible scientific journey.**

Mentoring and collaborating with younger colleagues has been one of the most fulfilling parts of my PhD. I am especially grateful to **Prabhleen** (for her help with repair assays), **Prosad and Vishal** (for bioinformatics analysis), and **Mahima Dey** (for her generous support in so many ways). Thank you, **Lalit**, for your constant assistance that kept my experiments on track. And to the ever-energetic juniors—**Priya, Talat, Manisha, Kedar, Inderpreet, Abdul, Shubham, Harshit, Aarushi, Kanikah, Afzal, Mahima, Harita, Sajal, Susi, and Punam**—thank you for your

dedication and friendship.

**This list is not complete without the mention of my batchmates especially Bani, Vicky and Aakanksha, Dheeraj, Muneesh, Himani, Vandana, Babita and Deepanshi who helped me whenever I needed.**

It would be unjust if I do not acknowledge abcicon, BioGenesis, Centre for infectious disease research (CIDR) and Indian institute of Science (IISc), Department of Biochemistry of Shaheed Rajguru College of Applied sciences for Women, Department of Microbiology of Bhaskaracharya College of Applied Sciences, University of Delhi for giving me platform to talk about my PhD work.

My sincere gratitude for the support of the hostel and canteen staff, sanitation staff, computers and technical staff, HR and PME staff, especially Nikhil ji and Jitendra ji, accounts and AcSIR staff to finish my doctoral scientific venture.

I acknowledge **Director IGIB and CSIR for research infrastructures, AcSIR for PhD degree, CSIR and ICMR for fellowships which enable me for timely completion.**

**I consider myself lucky and blessed to have Yasmeen, Dr.Surbhi, Vikram, Aakanksha Sharma and Bani Jolly as friends which I made in IGIB that helped me sustain my COVID times better.**

**I feel blessed to work, laugh, and grow with Pratibha Chaurasia who is not only my ex-roommate but she becomes a family to me in New Delhi. She has not just witnessed all my ups and downs of this journey but made it memorable and helping me to become more stronger in life. She is one of the major contributors for completing this thesis along with these special bunch of people- MOMENTORS: Varsha Chokker, Shikha Rani, Uma Sharma, Himanshu, Jatin, Prashant, Mandeep, Christeenaa and Nitika.**

I would especially like to acknowledge a beautiful couple **Dr. Rajeshwari Tiwari and Aman Mewara** for always being an amazing host, my go to people, my gol-gappe partners and many other partners in crime which I don't bother to share. Now it's time to mention the cutest and adorable person **Shree Aman Mewara** who became my stress buster specially during my thesis submission

time.

To anyone whose name I may have inadvertently missed, please know your contributions have not gone unnoticed—I am sincerely grateful to you all.

**Finally, my everlasting gratitude to my family for believing in me. This PhD is the ripe fruit of the seeds of sacrifices my family has sown over a long period. I dedicate this doctorate to my mom, dad and annoying brother Shrey for enduring all the challenges life has thrown at us and not letting an inch of it affect me. My mother, Mrs. Neeru Phutela, has always been more understanding and enthusiastic than I am about my PhD completion. She is the one who has always taught me to maintain a healthy work-life balance and to take care of my mental health, which is what I believe to be the most essential things in life. I am honoured to be called your daughter, Maa. Thank you for being who you are. My Dad, Mr. Ajay Kumar Phutela, no words of mine can express the things I really feel for you Papa. My father, Mr. Ajay Kumar Phutela, is one of the main contributors to my role and my inspiration. Your commitment and love for your work to this day inspire me more than words can say. I appreciate you being who you are and being there for the rest of us, Dad! It's time to acknowledge my brother Shrey Phutela, who intends to kill me if I miss this. Thank you very much for always providing me twice as much money as I ask for at the end of each month. You have consistently stood by me through thick and thin. My sanity has been preserved by your unwavering love and support throughout- I love you.**

**Who says dreams must pause after marriage? I stand as living proof that they only grow brighter, thanks to the unwavering love and encouragement of my incredible in-laws, Mrs. Anjana Kamra and Mr. Anil Kamra, and above all, my partner in every sense, Aayush. With their constant support, I've found the freedom to chase my aspirations while embracing life's joyful moments to the fullest. The beautiful family Aayush brought into my life has been a source of endless motivation, reminding me to aim higher, laugh louder, and dream bigger. And knowing Aayush, I'm sure he'll keep playfully pushing me to be my best, as only he can.**

**I feel so fortunate to share this journey with the most vibrant, loving people who fill my world with happiness—Arpita, Aakanksha, Akshit, Anika, Jiya, Manu Bhaiya, my ever-graceful Pari Bhabhi, and their adorable bundle of joy Advika , Ishika, Ankushi, Rabia, Yugu, Vansh,**

all my Bhua's and Massi's, Sabia Di, Sahil Jiju with their little star Aayan, Sameeksha, Mainak, Plash, Ankit—the list of these precious souls goes on and on.

I couldn't feel luckier to have not just one, but two incredible families to celebrate this milestone with me. This chapter of my life is richer, fuller, and infinitely more meaningful because of each one of you.

This was a poignant reminder of the power of gratitude and the importance of acknowledging those who have helped me shape my path.

I conclude this acknowledgement of my PhD journey by *saying “Remember, you have been criticizing yourself for years and it hasn’t worked. Try approving of yourself and see what happens” and this Happened to me .....*

## Vernacular abstract

छात्र का नाम: रिदम फुटेला

प्रस्तुति का वर्ष: 2024

पर्यवेक्षक का नाम: डॉ. देबज्योति चक्रवर्ती

सीएसआईआर लैब: इंस्टीट्यूट ऑफ जीनोमिक्स एंड इंटीग्रेटिव बायोलॉजी

पंजीकरण संख्या: 10BB18A02011

अध्ययन संकाय: जैविक विज्ञान

थीसिस का शीर्षक: कैस ९ मध्यस्थता दरार से प्रेरित डीएनए मरम्मत तंत्र की व्यवस्थित समझ।

क्लस्टर्ड रेगुलरली इंटरस्पेस्ड शॉर्ट पालिंड्रोमिक रिपीट कैस प्रणाली का उपयोग अब कई नैदानिक परीक्षणों में चिकित्सीय जीनोम संपादन के लिए किया गया है। क्लस्टर्ड रेगुलरली इंटरस्पेस्ड शॉर्ट पालिंड्रोमिक रिपीट कैस द्वारा प्रेरित डीएनए ब्रेक की मरम्मत कोशिका के अंदर विभिन्न मरम्मत प्रोटीनों द्वारा की जाती है, जिनमें से एक क्लासिकल नॉन-होमोलॉग्स एंड-ज्वाइनिंग (सीएनएचजे) है जो यादच्छिक सम्मिलन या विलोपन (इंडल्स) द्वारा जीन व्यवधान की ओर ले जाता है। वैकल्पिक रूप से, सटीक डीएनए संपादन लक्ष्य होमोलॉजी वाले एक बहिर्जात डीएनए टेम्पलेट प्रदान करके होता है जो होमोलॉजी डायरेक्टेड रिपेयर (एचडीआर) के माध्यम से वांछित डीएनए परिवर्तन को एन्कोड करता है। कोशिकाओं में दोनों मरम्मत मार्ग एक-दूसरे के साथ प्रतिस्पर्धा करते हैं, लेकिन एनएचर्जे एचडीआर की तुलना में अधिक कुशल है क्योंकि एचडीआर कोशिका चक्र चरण और लक्ष्य जीनोमिक स्थान पर निर्भर करता है। अनुसंधान जीनोमिक संशोधनों को बेहतर बनाने के लिए एचडीआर के लिए सेलुलर मरम्मत तंत्र को प्रभावित करने पर केंद्रित है, जिसमें कैस वेरिएंट उच्च एचआर आवृत्तियों को दिखाते हैं, लेकिन इन पूर्वाग्रहों के पीछे के तंत्र अस्पष्ट हैं। क्लस्टर्ड रेगुलरली इंटरस्पेस्ड शॉर्ट पालिंड्रोमिक रिपीट कैस सटीक डीएनए संपादन की क्षमता का पूरी तरह से उपयोग करने के लिए, डीएनए घावों की प्रकृति और मरम्मत मार्गों का अध्ययन करना महत्वपूर्ण है। हाल के अवलोकन एनएचर्जे की तुलना में उच्च होमोलॉजी निर्देशित डीएनए मरम्मत (एचडीआर) दर पैदा करने के लिए एफएनसीएएस९ और इसके इंजीनियर वेरिएंट (इएन एफ एन कैस ९) की व्यवहार्यता दिखाते हैं। यह ध्यान में रखते हुए कि एफ एन कैस ९ और इसका इंजीनियर संस्करण दरार के बाद 2-3 बेस कंपिट डबल स्ट्रैंड ब्रेक उत्पन्न करता है (कैस १२ की पिछली रिपोर्टों के समान), हमारा लक्ष्य डीएनए मरम्मत प्रोटीन की भर्ती गतिशीलता को समझना है जो कुंद-समाप्त दरार पर उत्पन्न होने वाले से भिन्न हो सकते हैं। हमने यह जांचने के लिए आनुवंशिक गड़बड़ी, अनुक्रमण और जैव रासायनिक परीक्षणों के संयोजन का उपयोग किया है कि क्या कैस ९-प्रेरित डीएनए घावों की प्रकृति मरम्मत के परिणामों को प्रभावित करती है। मरम्मत प्रोटीओम की पहचान करने से हमें उन कारकों की सराहना करने में मदद मिल सकती है जो एफ एन कैस ९-प्रेरित डबल स्ट्रैंड ब्रेक के जवाब में डीएनए मरम्मत मार्ग की पसंद को प्रभावित कर सकते हैं। कुमिलाकर, हम एचडीआर आधारित चिकित्सीय परिणामों में सुधार की संभावनाओं की जांच कर रहे हैं और इस अत्यधिक सटीक डीएनए पूछताछ प्रोटीन की क्रिया के तंत्र को समझ रहे हैं।

## Abbreviations

<b>Abbreviations</b>	<b>Definition</b>
<b>alt-NHEJ</b>	Alternative non-homologous end joining
<b>A-EJ</b>	Alternative non-homologous end joining
<b>Amp</b>	Ampicillin
<b>BCA</b>	Bicinchoninic Acid
<b>BFP</b>	Blue fluorescent protein
<b>B-ME</b>	$\beta$ -mercaptoethanol
<b>BSA</b>	Bovine Serum Albumin
<b>Cas</b>	CRISPR associated protein
<b>Cascade</b>	CRISPR-associated complex for antiviral defense.
<b>c-MYC</b>	Cellular myelocytomatosis
<b>cNHEJ</b>	Classical non-homologous end joining
<b>Cpf1</b>	CRISPR from Prevotella and Francisella 1
<b>CRISPRi</b>	Clustered Regularly Interspaced Short Palindromic Repeats Cas9 interference
<b>CRISPR</b>	Clustered Regularly Interspaced Short Palindromic Repeats
<b>CRISPRDx</b>	CRISPR Diagnostics
<b>crRNA</b>	CRISPR RNA
<b>crRNP</b>	CRISPR ribonucleoprotein
<b>DAPI</b>	4',6-diamidino-2-phenylindole
<b>dCas9</b>	Dead Cas9
<b>DMEM</b>	Dulbecco's modified Eagle Media
<b>DMSO</b>	Dimethyl sulfoxide
<b>DNA</b>	Deoxyribonucleic acid
<b>DPBS</b>	Dulbecco's phosphate-buffered saline
<b>DSB</b>	Double Strand Break
<b>DSTR</b>	single-stranded DNA donor templated repair
<b>dsDNA</b>	Double strand deoxyribonucleotide
<b>dsODN</b>	Double stranded oligodeoxynucleotide
<b>DTT</b>	Dithiothreitol
<b>EB</b>	Elution buffer
<b>EDTA</b>	Ethylene diamine tetra acetic acid
<b>EMSA</b>	Electrophoretic mobility shift
<b>EMX1</b>	Empty Spiracles Homeobox 1
<b>enFn</b>	Engineered <i>Francisella novicida</i>
<b>ESCs</b>	Embryonic stem cells
<b>EtBr</b>	Ethidium Bromide
<b>FACS</b>	Fluorescence-activated cell sorting

<b>FANCF</b>	Fanconi anemia complementation group F
<b>FASN</b>	Fatty acid synthase
<b>FBS</b>	Fetal Bovine Serum
<b>FELUDA</b>	FnCas9 Editor Linked Uniform Detection Assay
<b>FITC</b>	Fluorescein isothiocyanate
<b>FnCas9</b>	<i>Francisella novicida</i> Cas9
<b>gDNA</b>	Genomic deoxyribonucleotide
<b>GFP</b>	Green fluorescence protein
<b>GRCh38</b>	Genome Reference Consortium Human Build 38
<b>GUIDE-Seq</b>	Genome-wide unbiased identification of double stranded breaks enabled by sequencing
<b>HBB</b>	Hemoglobin Subunit Beta
<b>HDR</b>	Homology Directed Repair
<b>HEK</b>	Human Embryonic Kidney
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<b>HPLC</b>	High-performance liquid chromatography
<b>HR</b>	Homologous recombination
<b>HRP</b>	Horseradish peroxidase
<b>HS</b>	High sensitivity
<b>IF</b>	Immunofluorescence
<b>IP</b>	Immunoprecipitation
<b>iPSC</b>	Induced pluripotent stem cells
<b>IPTG</b>	Isopropyl-β-D-thiogalactopyranosid
<b>IVC</b>	in vitro cleavage
<b>IVT</b>	in vitro transcription
<b>KI</b>	Knock In
<b>KO</b>	Knockout
<b>KRAB</b>	Krüppel-associated box
<b>LIC</b>	Ligation independent cloning
<b>LMNA</b>	Lamin
<b>mESCs</b>	Mouse embryonic stem cells
<b>mL</b>	MilliLiter
<b>MM</b>	Mismatches
<b> mM</b>	Millimolar
<b>MMEJ</b>	Micro-mediated end joining
<b>MST</b>	Microscale Thermophoresis
<b>Nc</b>	Negative control
<b>NCBI</b>	National Center for Biotechnology Information
<b>NEAA</b>	Non-Essential Amino Acids
<b>ng</b>	Nanogram
<b>NHEJ</b>	Non-Homologous End Joining

<b>Nt DNA</b>	non-Target DNA
<b>nM</b>	Nanomolar
<b>NmCas9</b>	<i>Neisseria meningitidis</i> Cas9
<b>NT</b>	Nucleotide
<b>OCT-4</b>	Octamer-binding transcription factor 4
<b>OT</b>	Off-target
<b>PAGE</b>	Polyacrylamide gel electrophoresis
<b>PAM</b>	Protospacer Adjacent Motif
<b>PBMC</b>	Peripheral blood mononuclear cells
<b>PBS</b>	Phosphate-buffered saline
<b>PBST</b>	Phosphate buffered saline with tween detergent
<b>Pc</b>	Positive control
<b>PCR</b>	Polymerase Chain Reaction
<b>PFA</b>	Paraformaldehyde
<b>PIC</b>	Protease inhibitor cocktail
<b>POC</b>	Point of care
<b>PUC-119</b>	Plasmid University of California 119
<b>qPCR</b>	Quantitative Polymerase Chain Reaction
<b>RAY</b>	Rapid Variant AssaY
<b>RNA</b>	Ribonucleic acid
<b>RNP</b>	Ribonucleoprotein
<b>RT</b>	Room temperature
<b>RT-PCR</b>	Reverse transcriptase- Polymerase chain reaction
<b>SCA</b>	Sickle Cell Anaemia
<b>SCD</b>	Sickle Cell Disease
<b>SCF</b>	Stem Cell Factor
<b>SD</b>	Standard Deviation
<b>SDS</b>	Sodium dodecyl sulfate
<b>SEM</b>	Standard Error of Mean
<b>sgRNA</b>	Single-guide RNA
<b>SHERLOCK</b>	Specific High-sensitivity Enzymatic Reporter un-LOCK- ing
<b>SNP</b>	Single nucleotide polymorphism
<b>SNV</b>	Single nucleotide variations
<b>SpCas9</b>	Streptococcus pyogenes Cas9
<b>SSA</b>	Single strand annealing
<b>ssDNA</b>	single-stranded deoxyribonucleotide
<b>ssODN</b>	single-stranded oligodeoxynucleotide
<b>SSTR</b>	Single stranded DNA donor templated repair
<b>STR</b>	Short tandem repeat
<b>sx-Grna</b>	Super extended guide-ribonucleotide
<b>TALEN</b>	Transcription activator-like effector nucleases

<b>TDNA</b>	Target DNA
<b>TOPSE</b>	True Outcome Predicted via Strip Evaluation
<b>tracrRNA</b>	Trans-acting CRISPR RNA
<b>uL</b>	Microliter
<b>USER</b>	Uracil-Specific Excision Reagent
<b>UV B</b>	Ultraviolet radiation B
<b>VEGFA</b>	Vascular endothelial growth factor
<b>WT</b>	Wild type
<b>x-gRNA</b>	Extended guide-ribonucleotide
<b>ZFN</b>	Zinc Finger Nuclease

## List of Figures

- Figure 1.1 A schematic of CRISPR-Cas adaptation and defense mechanism.
- Figure 1.2: Classification and modular organization of CRISPR-Cas systems.
- Figure 1.3: The orchestration of repair pathways induced by Cas9 cleavage in mammalian cells.
- Figure 1.4: Principle of CRISPR-Cas9 based genome editing and outcomes in mammalian cells.
- Figure 1.5 Schematic summarizing the strategies for increasing the HDR/NHEJ ratio in cells.
- Figure 1.6 Indel events (shown as black dots and expressed as a percentage) discovered using amplicon sequencing via plasmid transfections in HEK293T.
- Figure 1.7 FnCas9 and its engineered variants can distinguish SNVs via lateral flow assay.
- Figure 1.8 Bar plot showing the HDR frequencies mediated by FnCas9 and en1FnCas9.
- Figure 1.9 Models depicting the engagement of repair proteins depending upon the ends of the DSBs.
- Figure 2.1 FnCas9 and its engineered variants create 2-3 base staggered DNA cleavage under in-vitro conditions.
- Figure 2.2 Experimental paradigm of template dependent polymerase filling in case of Cas9 induced DNA repair.
- Figure 2.3 Workflow for investigating en1FnCas9 mediated influenced insertions.
- Figure 2.4 Experimental paradigm of processing of in-vitro generated blunt and staggered DNA products by Cas9.
- Figure 2.5 Repair accuracy of blunt DNA ends is more than staggered DNA ends in bacterial cells.
- Figure 2.6 Repair accuracy of blunt DNA ends is less than staggered DNA ends in mammalian cells.
- Figure 3.1 Generation of stable dspCas9 KRAB expressing HEK293T cells.
- Figure 3.2 Experimental workflow for investigating repair dynamics induced upon SpCas9 and en1FnCas9 cleavage.
- Figure 3.3 Characterisation of CRISPRi library (AX227) inhibiting DNA repair and associated processes.
- Figure 3.4 Microscopic images of Untransduced and transduced BFP positive AX227 library in dspCas9 KRAB mcherry positive HEK293T cells.
- Figure 3.5 Single cell RNA sequencing of cells undergone for repairing staggered and blunt ended DNA products.
- Figure 3.6 Pipeline for single cell RNA processing via BD Rhapsody platform.

- Figure 3.7 Quality control, Normalisation and Clustering of Single Cell RNA sequencing.
- Figure 3.8 Differential Expression Analysis of blunt and staggered ends DNA repair genes.
- Figure 3.9 Schematic summarizing the processing of staggered DSB induction and its repair.
- Figure 4.1 Experimental paradigm showing different lengths and types of donor templates that repair the Cas9 induced DNA breaks via HDR.
- Figure 4.2 Experimental paradigm showing the Cas9 expressing stable cells undergoing HDR repair pathways in presence of longer DNA template.
- Figure 4.3 Generation of stable Cas9 expressing HEK293T cells.
- Figure 4.4 Cas9 expressing stable HEK293T cells are not prone to apoptosis.
- Figure 4.5 Frequency of HDR is higher during en1FnCas9 mediated editing.
- Figure 4.6 Experimental paradigm of investigating HDR/NHEJ in HEK293T cells.
- Figure 4.7 Stacked Bar plots showing the HDR and indel editing percentage in HEK293T cells.
- Figure 4.8 Bar plots showing the HDR/Indel ratio in HEK293T cells using ssODN and dsODN templates incorporated at targets (FASN, LMNA, HBB).
- Figure 4.9 Integrative Genomics Viewer (IGV) screenshot showing the aligned reads to the HBB region (5Kb) detected by nanopore sequencing.
- Figure 4.10 Experimental paradigm of stem cells undergoing HDR in presence of different donor types.
- Figure 4.11 Cellular editing of stem cells by different Cas9 targeting OCT4 locus.
- Figure 4.12 Bar plots showing the HDR/Indel ratio in HEScs cells using ssODN and dsODN templates incorporated at OCT4 locus.
- Figure 4.13 Experimental paradigm for investigating editing outcomes in mouse embryos.
- Figure 4.14 Editing in mouse zygotes by en1 FnCas9 targeting OCT4 locus.
- Figure 4.15 Genome wide Off Target analysis by digenome sequencing.
- Figure 5.1 Schematic summarizing this thesis work.
- Figure 5.2 Scission based targeting for correcting pathogenic deletions using CRISPR-Cas9 system.

## List of Tables

- Table A: List of the oligos required for amplifying the IVC templates from genomic loci and the CRISPR guide RNA.
- Table B: List of the oligos used for constructing and validating Cas9 stable cell lines.
- Table C: List of the oligos used for cloning the guide RNA sequences in mammalian expressing Cas9 plasmids.
- Table D: List of the oligos used for amplifying the edited loci (HDR/Indels) via high throughput sequencing.
- Table E: List of the oligos used for amplifying the edited loci (/Indels) for high throughput long range nanopore sequencing.
- Table F: List of the oligos used for making the ssODN and dsODN, phosphorothioate modified at 5' and 3' ends and introducing AAA as a signature sequence.
- Table G: List of the oligos used for making and validating CRISPRi HEK293T cells.
- Table H: List of the genes regulating staggered and blunt ended repair from the single cell RNA sequencing dataset.

## **Chapter 1: Overview and principle of CRISPR-Cas system**

# Introduction

## 1.1 Overview of CRISPR Cas system

The clustered regularly interspaced short palindromic repeat (CRISPR), RNA guided Genome editing tool which is currently used to edit various organisms and cell lines. Initially, it was identified as mysterious DNA sequences in various bacteria like *Escherichia coli*, *Mycobacterium tuberculosis* and later extended to *Streptococcus pyogenes* (Ishino et al., 1987; PM et al., 1993; Hoe et al., 1999). Some years down the line, eminent scientists around the world have observed that these unique loci contain fragments of foreign DNA, which comprises a portion of the bacterial and archaeal immune system. The discovery of CRISPR loci in almost 40% of bacterial genomes and 90% of archaeal genomes have opened up new avenues of research and shed light on their crucial role of the prokaryotic immune system (Sorek et al., 2008). These CRISPR loci are composed of viral DNA fragments called "spacers," typically ranging from 17 to 84 bases in length. These spacers are separated by short palindromic repeats that span around 23 to 48 bases (Papkov et al., 2016). These repeats are organized into clusters within intergenic regions, forming a collection of potentially harmful genetic information. The identification of viral DNA fragments and the mechanism of targeting foreign DNA represented a paradigm shift, offering potential applications in genomic editing (Makarova et al., 2011; Marraffini et al., 2008). This discovery underscored the potential of employing this system for editing genomes in laboratory settings (Abudayyeh et al., 2016; Abudayyeh et al., 2017).

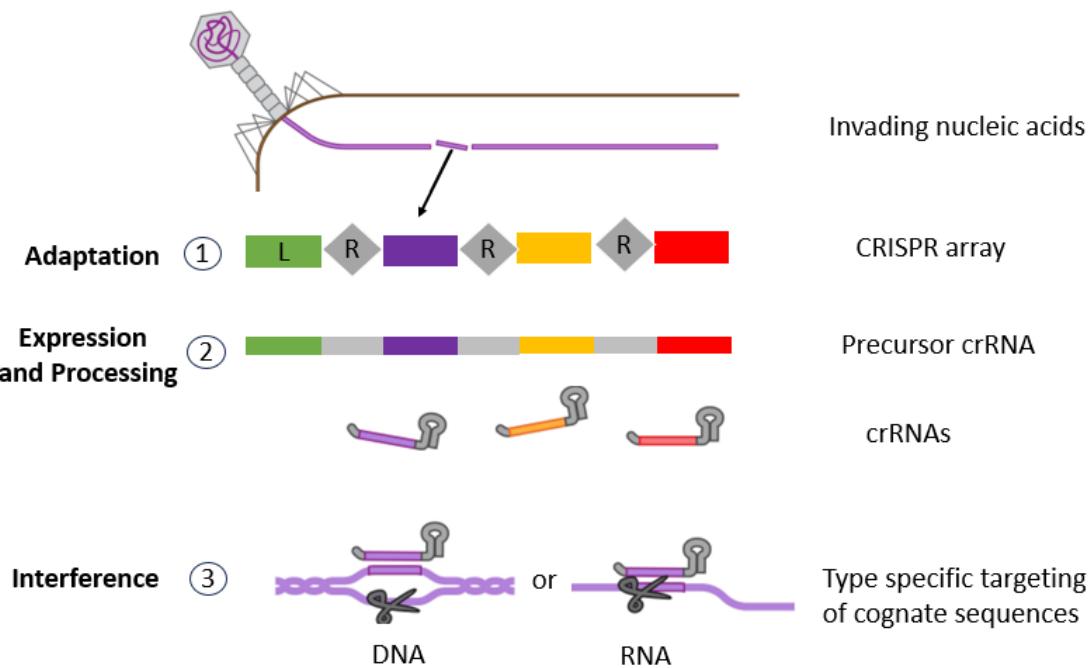
The CRISPR Cas systems naturally function as an immune mechanism in archaea and bacteria. This system shows a remarkable diversity of Cas protein sequences, genetic compositions and architectures of the genomic loci alike other biological Défense mechanisms in prokaryotes. CRISPR immune response mounts in three discrete yet intertwined stages such as adaptation, expression and interference.

During the Adaptation stage, a distinct complex of Cas proteins binds to the target DNA after recognizing a short sequence known as the protospacer-adjacent motif (PAM). At this adaptation phase, foreign portion of DNA fragments (protospacers) becomes spacer which is recognised and cleaved in the target DNA by the Cas complex. There are alternative adaptive stages employed by some CRISPR–Cas systems in which the spacer is acquired from RNA by a reverse transcriptase, which is encoded at the *CRISPR–cas* locus (Takeuchi et al., 2012; Koonin et al., 2022).

The next step involves the expression and maturation of CRISPR (cr) RNAs. During the expression stage, transcription of the CRISPR array to a single transcript, which is referred as pre-CRISPR RNA (pre-crRNA), is subsequently processed into mature CRISPR RNAs (crRNAs). Each crRNA contains the spacer sequence and parts of the flanking repeats.

In diverse CRISPR–Cas variants, either a multidomain Cas protein, a discrete subunit of a multiprotein Cas complex, a single or by non-Cas host RNases mediates this pre-crRNA processing.

The last stage is the **Interference** stage where the processed crRNA remains bound to the protein and guides this complex to identify and bind to the protospacer of the invading genome of a virus (DNA, RNA) or plasmid (DNA). This interaction triggers the nuclease component of the CRISPR effector complex to cleave the recognized target, thereby intervening the attack of the invaders such as virus or bacteriophages. (Figure 1.1).



**Figure 1.1 A schematic of CRISPR-Cas adaptation and Défense mechanism.**

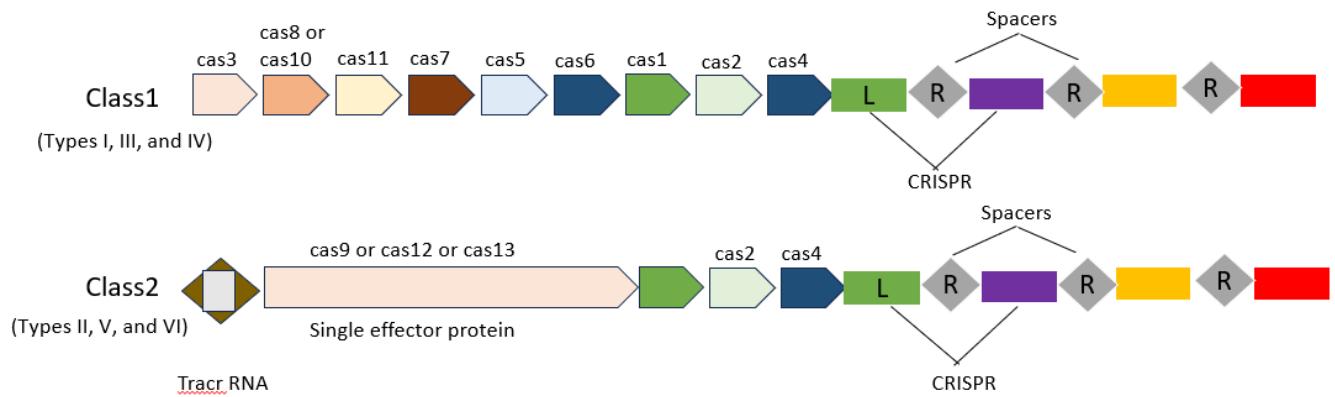
This system consists of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) array and CRISPR-associated (Cas) proteins which is encoded by cas genes. These are defined by three stages 1) Adaptation, a memory of preceding infections created by the incorporation of small foreign DNA elements into the leader (L) end of the CRISPR array, where they are reserved as spacers (coloured squares) between duplicated repeats (R). 2) Expression and biogenesis of

CRISPR-RNA (crRNA), the transcription and processing of the CRISPR array into small functional guide RNA sequences. 3) Interference, sequence-specific DNA or RNA binding and degradation of the targeting foreign invaders.

## 1.2 Classification of CRISPR system

Not all genes are common to all the CRISPR Cas systems, therefore necessitates its comprehensive phylogenetic classification in two categories Class 1 and Class 2. Class 1 systems demand for multi-subunit protein effector complexes that are composed of multiple Cas proteins and Class 2 systems involve a single, large, multidomain protein as the effector (Figure 1.2).

The two classes of CRISPR-Cas systems can be further sub-classified into six distinct types (I–VI) and this classification is contingent on cas gene signatures, cas operon organization, and phylogenetic analyses of conserved Cas proteins. Additionally, Class 1 and Class 2, are classified into three types. Class 1 includes types I, III, and IV, whereas class 2 encircles types II, V, and VI. Each class type is uniquely characterized by specific signature proteins formulated as operon architectures. Each class type is composed of multiple subtypes which are distinguished by locus organization and often accompanied by specific Cas proteins (Makarova et al., 2020).



**Figure 1.2 Modular organization and classification of CRISPR-Cas systems.** The general architectures of two classes of CRISPR-Cas systems. Cas genes are indicated as arrows and the homologous genes are also shown in the same colour.

The major difference between the classes of CRISPR-Cas systems lies in the maturation process of

pre-crRNA forms. In class 1 systems, a specialized multiprotein Cas complex processes the crRNAs into their mature form and is commonly referred as Cascade (CRISPR associated complex for antiviral Defense). The interaction of this complex with the pre-crRNA is aided by Cas proteins like Cas6 or Cas5. In contrast to type I and III CRISPR Cas systems, that accounts for majority of CRISPR loci and involves number of Cas proteins for the processing of crRNA, type II CRISPR Cas systems, completely utilizes different mechanisms (Makarova et al., 2020).

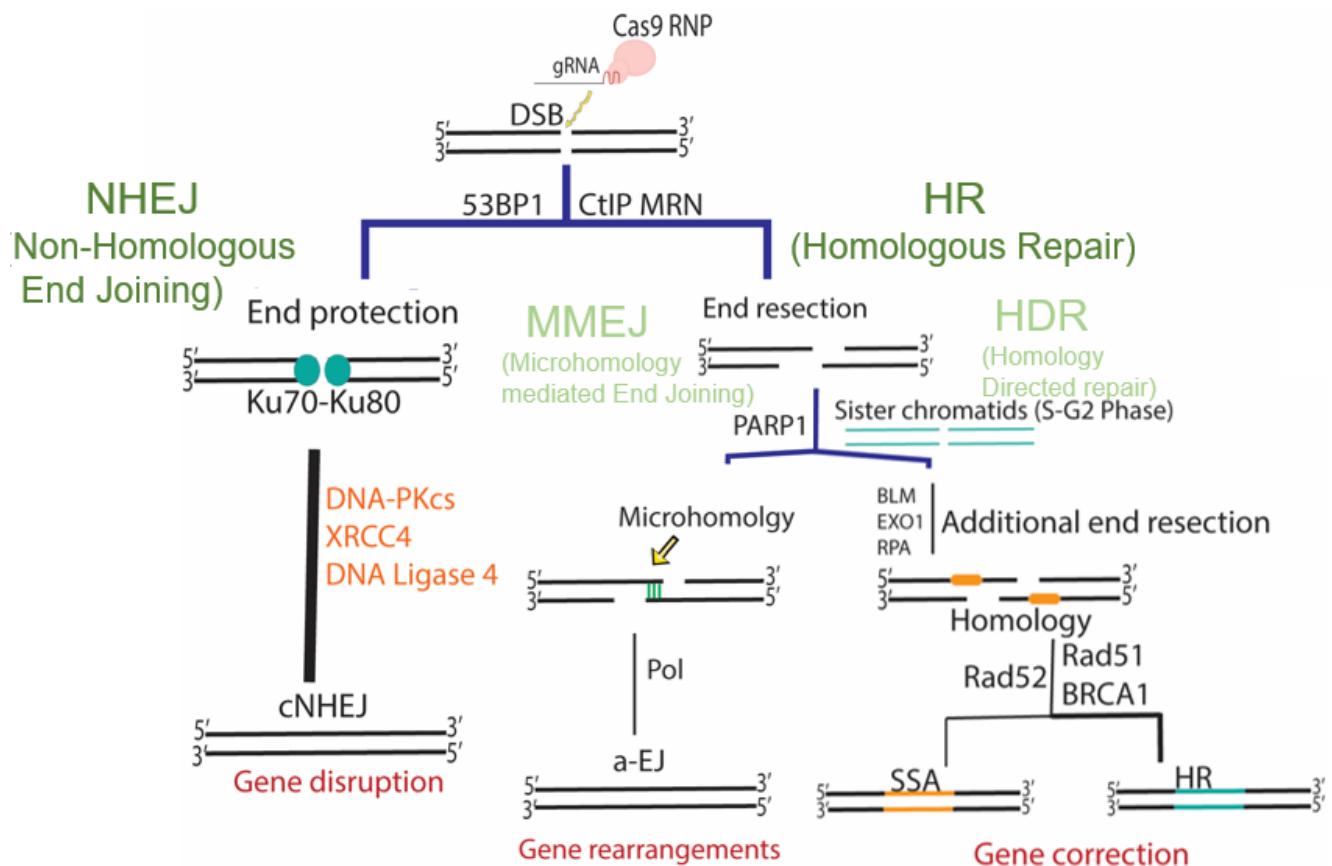
In type II systems, RNase III which is the host's own enzyme is deployed along with the trans-acting CRISPR RNA (tracrRNA). Hence, the processing of pre-crRNA and target cleavage is effectively executed by effector Cas9 DNA endonuclease, different from the type I and type III systems (Deltcheva et al., 2011). This endonuclease recognizes and cuts the target and non-target DNA strand exclusively by distinct nuclease domains like HNH or RuvC respectively (Jinek et al., 2012; Nishimasu et al., 2014). During this, the crRNA forms a dual-RNA hybrid structure with the tracrRNA, which is the guiding force for the Cas9 RNP to the target DNA sequences (adjacent to PAM) that is complementary with the cRNA region in the guide. It is especially seen in type II systems that the presence of tracrRNA is crucial for crRNA maturation (Gupta et al., 2014). To maintain the simplicity of the CRISPR Cas9 system, a hybrid single-guide RNA that is known as sgRNA has been created (combination of crRNA and tracrRNA) but retained its specificity to the target DNA. It increases the robustness of the type II CRISPR Cas9 system as it targets any desired DNA sequence within the genome very easily compared to earlier genome engineering tools.

### 1.3 Repair pathways induced by CRISPR Cas DNA cleavage

Organisms possess billions of bases of either DNA or RNA within their genomes, hence encompassing a wide range of life forms and the arrangement of these bases are predetermined. CRISPR which is an RNA guided genome editing tool, currently used to edit the genome of various organisms and cell lines. The concept of precisely targeting these DNA bases holds extraordinary potential across numerous fields like agriculture, disease therapy, diagnostics, screening drugs, in short from molecular biology to biomedicine and beyond (Chapman et al., 2012; Mali et al., 2013).

Most canonical type II Cas9 which is used for genome editing purposes is SpCas9. It recognises and binds to the target DNA which is 20 nucleotides long through Watson–Crick base pairing in a sequence-specific manner called spacer, lies near the protospacer adjacent motif (PAM- NGG). It cleaves the target via HNH nuclease domain and the non-target strand via RuvC nuclease domain

inducing a DNA double-strand break (DSB). Creation of a DSB in mammalian cells at a specific locus ignites the process of gene editing. These DNA lesions could be resolved by any of these four DNA repair pathways, one of the two major pathways are canonical NHEJ (c-NHEJ) and homology-directed repair (HDR with additional ones namely, alternative NHEJ (alt-NHEJ) or Micro-mediated end joining and single-strand annealing (SSA) (Yang et al, 2020) (Figure 1.3).



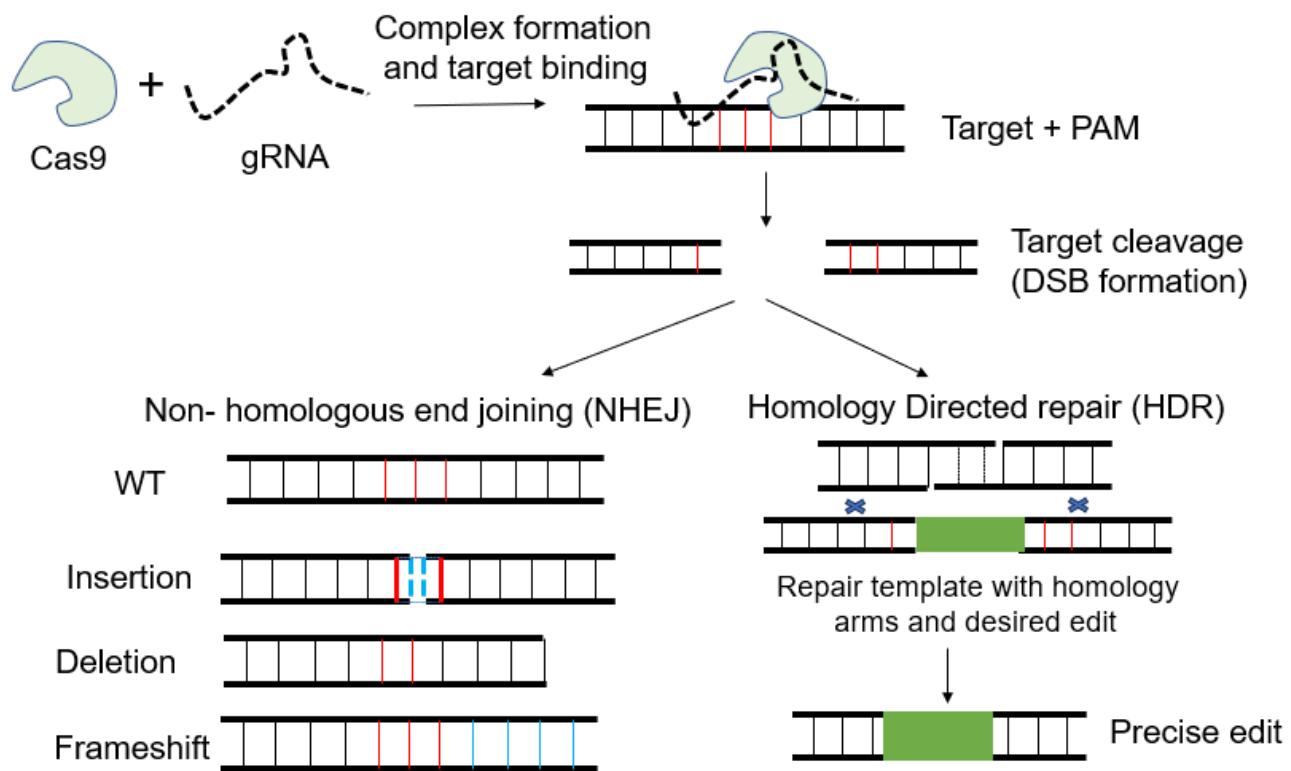
**Figure 1.3: The orchestration of repair pathways induced by Cas9 cleavage in mammalian cells.** (Left) The first choice of repairing these DSBs (double strands breaks) in mammalian cells is c-NHEJ which occurs during all the phases of the cell cycle. (Right) Preliminary end resection of these DSBs is necessary for its repair via other uncommon pathways like alt-NHEJ, SSA and HDR mechanisms. Resection is relatively short during alt-NHEJ but extended resection is favoured in HDR. SSA is the opportunistic repair pathway that functions only in the presence of ssDNA ends as exogenous templates.

The choice of DNA repair pathways, particularly between those engaged in HDR and NHEJ, will determine the consequences of genome editing outcomes. DSBs switch on most commonly cellular

DNA repair machinery which leads to random insertions or deletions (indels) therefore, gene disruption (Knock-out) through the c-NHEJ (Cong et al., 2013, Jinek et al., 2013).

Alternatively, a less commonly utilized cellular repair pathway HDR precisely edits the DNA (Knock-Ins) by introducing the desired DNA change in the exogenous DNA template and is flanked by the homologous region to the DSBs (Figure 1.4).

It offers the opportunity to introduce specific changes, such as substitutions and insertions of long stretches of DNA. This idea of programmable genome editing became tangible when it was discovered that inducing DSBs in the human cells triggers cellular repair processes, significantly increasing the frequency of alterations in the genome (Symington et al., 2011).



**Figure 1.4: Principle of CRISPR-Cas9 based genome editing and outcomes in mammalian cells.**

Cas9 in green colour binds with the gRNA and targets the dsDNA, once it cleaves the DNA the repair pathways that the cells take is NHEJ (Left) that could lead to Wild Type (WT), Insertion, Deletion and Frameshift. Less Prominent pathway is HDR that could lead to precise editing in cells (Right).

Amongst the NHEJ and HDR repair pathways, that competes with each other in cells, NHEJ won and becomes the more efficient one in most conditions depending upon cell type, state, cell cycle phase, and the genomic locus that is targeted.

## 1.4 Factors affecting DNA repair outcomes induced by CRISPR Cas9 cleavage

**1.4.1 Cell cycle:** In general, mammalian cells tend to use c-NHEJ throughout the cell cycle, although HR and SSA functions mainly during the S/G2 phase. c-NHEJ is a commonly used pathway for religating directly the broken DNA ends with minimal end processing, but is also inaccurate and often leads to new mutation (Lieber et al., 2010; Fell et al., 2015; Kraglund et al., 2016; Davis et al., 2014).

**1.4.2 Sequence of Target site:** Despite DNA editing outcomes considerably vary among different genomic sites but are not completely arbitrary. These are significantly affected by the nucleotide sequences adjacent to the Cas9 cutting site (Shen et al., 2018; Chakrabarti et al., 2019; Allen et al., 2018). For template-independent repair pathways, most common repair DNA outcomes are >3bp deletions which are the by-products of the MMEJ. Other DNA repair outcomes are short 1-2bp indels and its percentage is affected by the 4th nucleotide position upstream from the PAM (Brinkman et al., 2018; Shi et al., 2019).

**1.4.3 Structure of Chromatin:** The cleavage efficiency of Cas9 in eukaryotic cells is exceedingly dependent on accessible DNA which is bound by histones to form nucleosomes (Clouarie et al., 2019; Isaac et al., 2016). Heterochromatin compaction and transcriptionally active euchromatin impact the Cas9 induced DNA repair editing outcomes like HR, MMEJ, SSA and c-NHEJ respectively (Schep et al., 2021). The local structure of chromatin like histone posttranslational modifications also affect the recruitment of DNA repair proteins.

**1.4.4 Donor DNA templates:** CRISPR Cas9 induced template-dependent DNA repair pathways utilize exogenous donor template for precise edits. Donor DNA templates like single stranded DNA (ssDNA) or double stranded DNA (dsDNA) repair the DNA lesions via SSA and HR pathways respectively. Due to the presence of micro homologies in the donor DNA templates, mis-integration of events through the MMEJ leads to imprecise edits. Double-stranded DNA donor templated repair (DSTR) occurs mainly through HR and single-stranded DNA donor templated repair (SSTR) occurs mainly through SSA and SDSA. If one end is repaired through and other through c-NHEJ or MMEJ, this leads to asymmetric HDR. The DNA end which is complementary to the 3' end of ssDNA

templates is repaired through SSA/SDSA and repair of the other end is still unknown. Its processing through MMEJ generates asymmetric HDR (imprecise and precise edits) that displays a bias directionality with respect to the orientation of the ssDNA templates.

New strategies for suppressing the MMEJ tend to improve both DSTR and SSTR based HDR which is discussed in the later session (Sung et al., 2006; Bhargava et al., 2016; Yeh et al., 2019). Extensive research endeavours are directed towards influencing cellular repair mechanisms in favor of HDR to enhance the effectiveness of generating accurate genomic modifications. The ability to harness programmable genome editing opens up a world of possibilities for scientific exploration and practical applications across a diverse array of organisms. By precisely manipulating the genetic material of organisms, we unlock tremendous potential for advancements in various fields.

## 1.5 Regulation of the trade-off between NHEJ and HDR

Despite the uncovering of Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) has significantly improved the genome editing efficacy, but every new target gene editing site using these systems requires re-engineering of new protein sets, which restricts their broad term applications (Cox et al., 2015). Precise genome editing via CRISPR Cas9 nucleases is entering the realm of attaining desired and safe editing outcome hence resulting in its rapid adoption for therapeutics (Ware et al., 2018; Doudna et al., 2020; Barrangou et al., 2016). Accurate gene editing is accomplished via the HDR pathway in cells which is less active and only restricted to S/G2 phases. Therefore, this proposes that controlled repair outcomes of the DNA lesions could be achieved by manipulating the choice of DSB repair pathways. It involves the selection of the strategies which favors the increased frequency of Cas9-mediated HDR mechanisms.

**1.5.1 Suppression of crucial NHEJ factors:** Reports claimed that DSBs generated during S and G2 phases of cell cycle are resolved concurrently by c-NHEJ and HDR pathways. This advised that inhibition of factors that are essential for NHEJ results in channelisation of repair proteins toward the HDR pathway (Vartak et al., 2015). DNA ligase IV which plays a crucial role in sealing DNA breaks during the final step of c-NHEJ. Inhibition of DNA ligase IV by small molecule inhibitors like SCR7 is suggested as a potential strategy to inhibit NHEJ and enhance HDR, which is the need of the hour (Singh et al., 2015; Zhang et al., 2018; Ma et al., 2016; Hu et al., 2018; Li et al., 2017). Besides small molecule inhibitors, other methods like shRNA-mediated gene silencing and proteolytic degradation of DNA ligase IV through E1B55K and E4orf6-mediated ubiquitination have also been

shown to increase HDR efficiency in mammalian cells. Researchers have also targeted critical factors like Ku70 and Ku80 that are known to bind to DSBs, upstream of the NHEJ repair pathway via siRNA targeting that can enhance CRISPR/Cas9-mediated HDR (Li et al., 2018). Chemical Inhibition of DNA-PK, a protein involved in NHEJ, via NU7441 and KU-0060648, has been shown to increase HDR rates at DSBs induced by Cas9 in different cell types (Robert et al., 2015). Inhibiting NHEJ may lead to the accumulation of unrepaired DSBs in cells, potentially resulting in cell death or late embryonic lethality. Therefore, the safety of using these inhibitors *in vivo* would need careful evaluation (Frank et al., 1998).

**1.5.2 HDR favouring factors:** RAD51 plays a principal role in HDR and small molecules like RS-1 which are RAD51 agonists enhance its binding activity. Reports suggested treatment of RS-1 in various cell types, including HEK-293A cells, rabbit embryos, zebrafish embryos, and human pluripotent stem cells stimulated Cas9 mediated HDR (Pinder et al., 2015; Kurihara et al., 2020). Ectopically expressing hRAD52 has also improved CRISPR-mediated HDR, especially when using single-stranded donor oligonucleotides (ssODNs) as donor DNA templates (Paulsen et al., 2017). Fusion of Cas9 proteins with HDR promoting factors like N-terminal fragment of CtIP or other factors like RAD52 and MRE11 has also been reported to enhance the desired editing efficiency in human cells (Charpentier et al., 2018; Tran et al., 2019). Another approach involves the creation of chimeric Cas9 constructs in which the Cas9 is fused with a specific domain [like HSV-1 alkaline nuclease (UL12)] that has been used to recruit the HDR facilitating complexes like MRN to the DSBs (Reuven et al., 2019). There are some unknown mechanisms which involve the use of small molecules, such as Brefeldin and L755505 for increasing the HDR knock-in efficiency of large fragments as well as point mutations in some contexts (Yu et al., 2015).

**1.5.3 Commutation of RAD18 with 53BP1:** 53BP1 is the vital regulator for the repair choice in between NHEJ and HDR at the site of Double stranded breaks. It has been suggested that inhibiting 53BP1 can manipulate the repair choice towards HDR. Identification of Rad18 as a potent HDR enhancer by inhibiting the 53BP1 recruitment to DSBs (Jayavaradhan et al., 2019; Nambiar et al., 2019).

**1.5.4 Pairing of Cas9 Activity with HDR-Active Cell Cycle Phase:** There are strategies which aim to enhance the precise genome editing by offering better control over the specificity and timing of Cas9-mediated DNA repair described as follows:

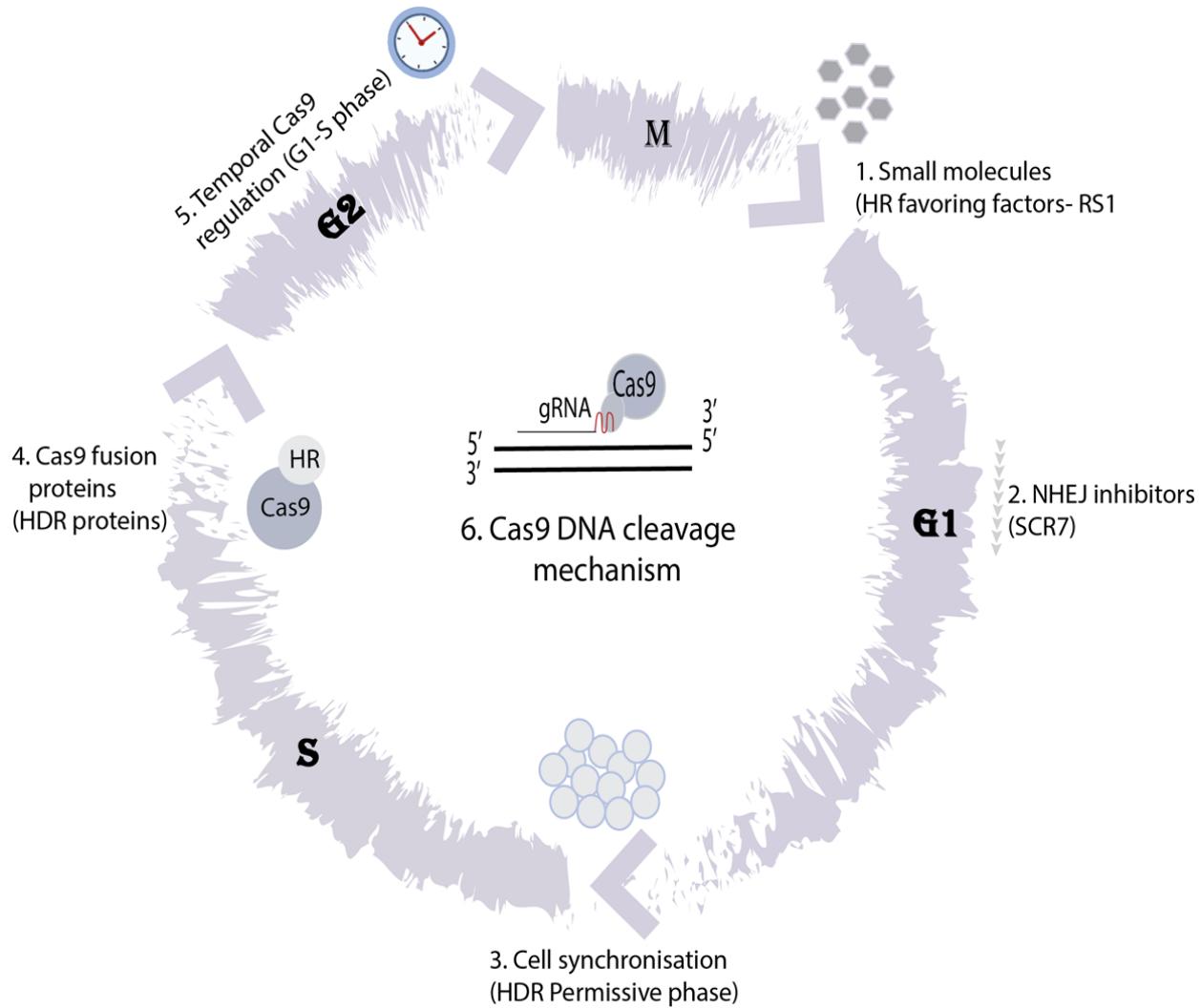
**Synchronizing cells to HDR-Permissive Cell Cycle Phase:** For improving cellular HDR efficiency, it is essential to target the cells via pharmacological means in an appropriate phase of cell cycle. Substances like Nocodazole (synchronize cells in G2 phase), ABT751 (synchronize induced pluripotent stem cells) and small molecules like RO-3306 (CDK1 inhibitor) and XL413 (targets initiation of DNA replication) have been shown to increase the efficiency of HDR in various contexts.

**Regulated Expression of Cas9:** One of the approaches to synchronize the generation of DSBs with the HDR-permissive cell cycle phase is by regulating the Cas9 expression. This is achieved by fusing Cas9 with the first 110 amino acids of Geminin, this resulted in the proteolytic destruction of Cas9 in the later cell cycle phases of M and G1. Combining the regulated Cas9 expression along with the cell synchronization molecules like RO-3306 has been significantly shown to improve the HDR/NHEJ ratio (Lin et al., 2014; Yang et al., 2016; Lomava et al., 2019; Wienert et al., 2020).

**1.5.5 Improving HDR by increasing the proximity of donor DNA template with CRISPR components:** This involves bringing the Cas9 and sgRNA which are the components of CRISPR System in close proximity of the HDR donor DNA template. Concepts like S1mplex, which uses sgRNA modified with streptavidin-binding aptamer and linked to a biotinylated single-stranded DNA (ssDNA) donor have shown improved HDR rates. In addition to this, Cas9 variant fused with avidin and biotin modified ssDNA was used as a bridge for increasing its affinity in HDR (Devkota et al., 2018; Carlson et al., 2017; Ma et al., 2017).

One of the approaches which eliminates the need to modify either sgRNA or donor ssDNA to enhance the HDR is tethering of ssDNA to the Cas9 RNP complex via a fused HUH Endonuclease. The donor types that can be used in these kinds of approaches are typically single-stranded oligodeoxynucleotides (ssODNs) or double-stranded DNA templates with long homology arms (1-2 kb). In some cases, linearised plasmids have been observed to perform better HDR editing than circular plasmids (Richardson et al., 2016; Song et al., 2017).

All these strategies solely aim to improve the efficiency of HDR by offering enough flexibility to choose donor molecules and reducing the need for extensive modifications that are costly and time-consuming (Figure 1.5).



**Figure 1.5 Schematic summarizing the strategies for increasing the HDR/NHEJ ratio in cells.** (1) Utilizing small molecules that favour Homology repair in cells like RS1. (2) Inhibition of NHEJ by using inhibitors like SCR7 in cells to increase the HDR frequency. (3) Synchronization of the cells in Homology directed repair permissive phase for increasing the HDR. (4) Fusing HDR proteins with Cas9 to increase the accessibility of these HDR proteins during Cas9 cleavage. (5) Temporal regulation of Cas9 during the G1-S phase of the cell cycle to increase the HDR. (6) Exploration of the Cas9 DNA cleavage mechanism in cells whether its blunt or staggered ends.

Various successful approaches have been developed for selectively disrupting NHEJ and boosting the HDR repair pathway. These strategies can vary depending on biological and experimental factors such as cell type, organism, genomic location, and experimental design. Manipulating repair pathways raises the need for contemplating safety aspects, such as inhibiting NHEJ which is a critical pathway for

genome stability and DSBs repair increases cellular toxicity both *in vitro* and *in vivo* settings. Moreover, chemical synchronization of cells also has limitations because of its potential toxicity. In summary, this emphasizes the challenges to enhance frequencies of HDR in genome editing, therefore, suggests that combining approaches and further understanding of repair pathways will be crucial for enhancing CRISPR/Cas9-mediated HDR.

Considering all these challenges and for simplifying the control of DNA repair outcomes, designing or exploring Cas9 nucleases that are inherently biased toward DSB repair by a specific pathway is desired. The central goal in this field is to develop a means for controlling the relative frequencies of specific repair pathways, particularly to make precise editing by HDR. This highlights the interest in developing or exploring Cas9 nucleases that have intrinsic repair biases to control the outcome of genome editing.

**1.5.6 Cas variants show higher HDR:** One of the strategies that have been recently explored in the field is how different Cas variants can show higher HDR frequencies. Engineered SpCas9 variants like SpCas9-HF1 and eCas9 particularly show potential in increasing HDR efficiency while minimizing off-target mutations. The reason for this is not known yet but it is proposed that the mutations for creating SpCas9 variants reduces its interactions with target DNA, and might also affect the DNA cleavage site by the RuvC lobe and can generate different DNA ends. It opens up the room to address whether the Cas9 based cleavage can affect the HDR by generating different DNA ends (Matsumoto et al, 2024).

Published study from our lab has revealed an increase in HDR frequencies by the Cas9 from *Francisella novicida* Cas9 (*FnCas9*) (Figure 1.10). The unique structural attributes and low tolerance for mismatches of *FnCas9* prompted an investigation into its DNA interaction properties and its role in genome editing. Unique features of *FnCas9* that distinguish it from other Cas9 enzymes is its high intrinsic specificity. It tolerates only a single mismatch at the PAM distal region of the sgRNA towards 5' position. In contrast, *SpCas9* can tolerate multiple mismatches primarily towards the non-seed region of sgRNA, leading to variable levels of off-target effects (Acharya et al., 2019).

We attempted a study to enhance the DNA cleavage efficiency of *FnCas9* by retaining its intrinsic specific property. This is done by introducing substitutions of the specific amino acids majorly in the WED-PI domain, aiming to establish new contacts with the PAM duplex DNA. It led to creation of 49 different variants of *FnCas9* based on its crystal structure. These variants are likely designed with specific modifications to improve their performance in terms of cleavage rate, without compromising the inherent specificity and fidelity of *FnCas9*. The goal is to enhance the speed of DNA cleavage without

making extensive structural changes to *FnCas9*. Through in vitro cleavage kinetics assays (IVC kinetics assay), the researchers discovered three engineered *FnCas9* variants **en1(E1369R)**, **en15(E1603H)** and **triple mutant en31 (E1369R/E1449H/G1243T)** that demonstrated around a two-fold higher rate of DNA cleavage compared to the wild-type *FnCas9* protein (Acharya et al., 2024). This indicates that these variants are more efficient in cleaving DNA even in an *in-vivo* scenario which is an encouraging result (Figure 1.6).

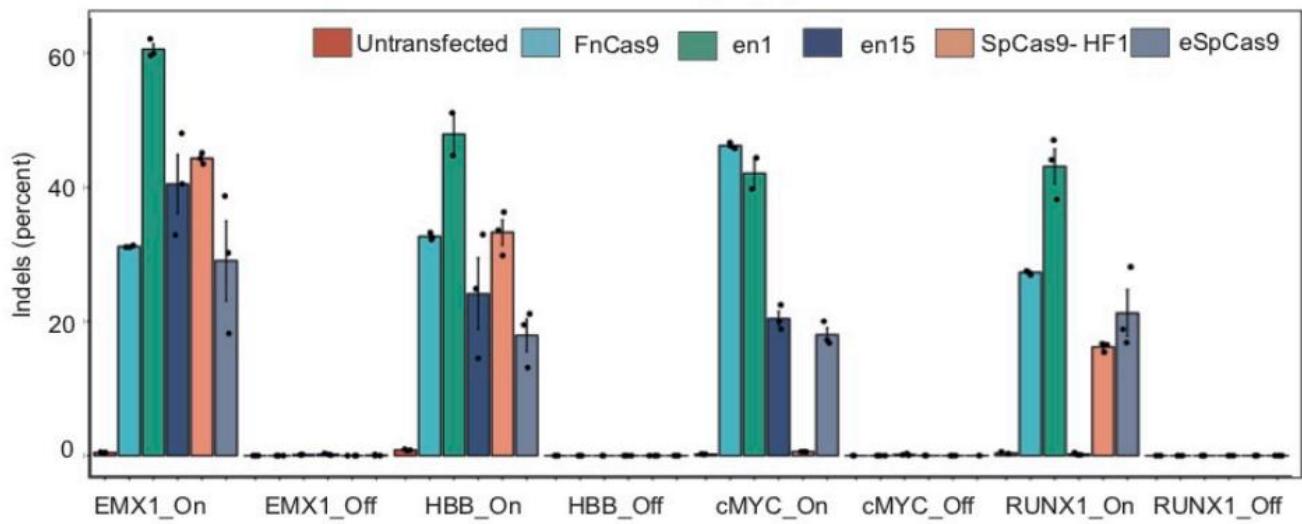
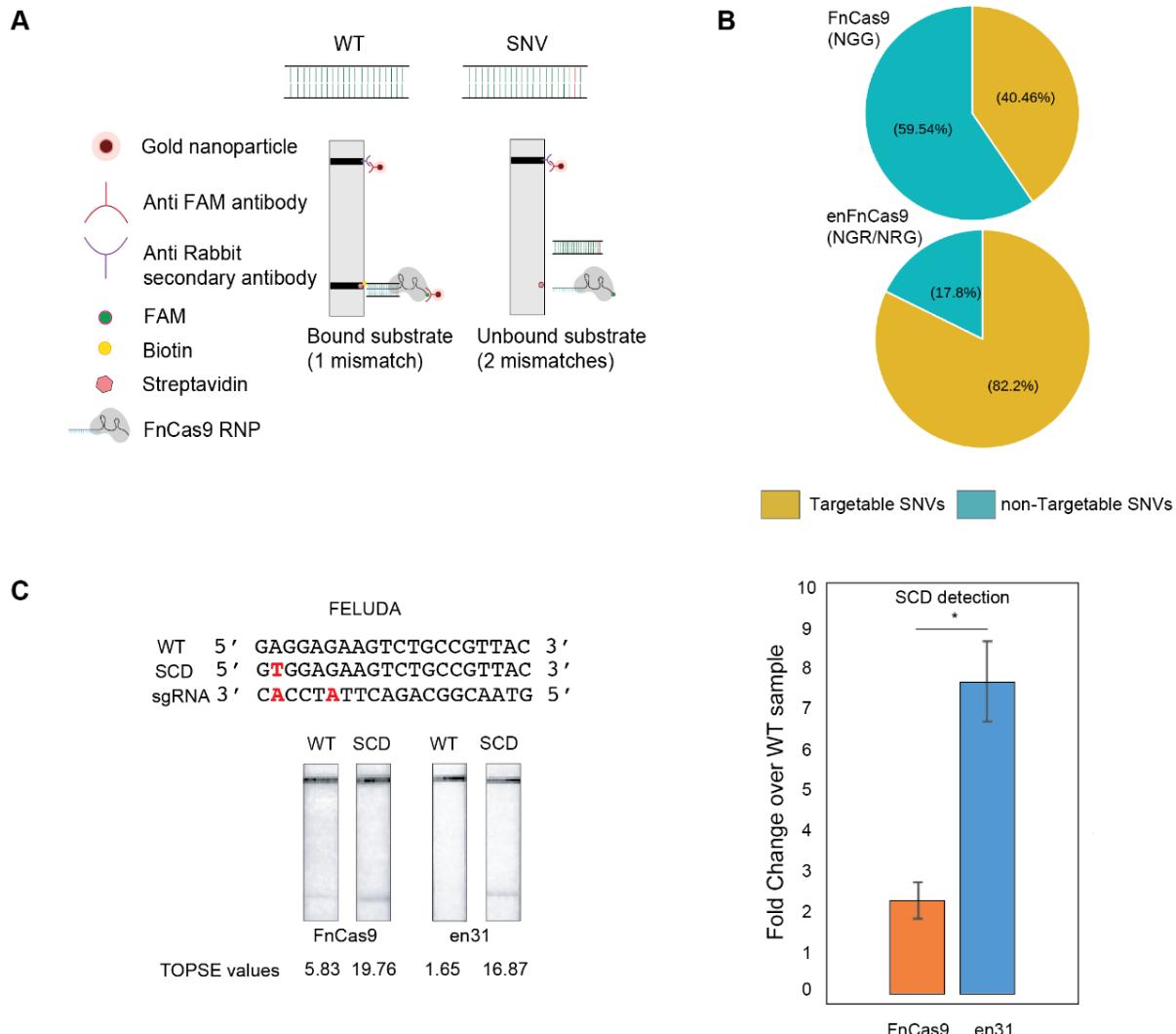


Figure 1.6 Indel events (shown as black dots and expressed as a percentage) discovered using amplicon sequencing via plasmid transfactions in HEK293T cells. *FnCas9*, *en1*, *en15*, *SpCas9-HF1*, and *eSpCas9* targeting the genes *EMX1*, *HBB*, *c-MYC*, and *RUNX1*. Untransfected cells are displayed as the control. Standard error mean for 3 separate experiments, excluding *cMYC* and *HBB* on-target of *en1FnCas9* is represented by error bars. (Figure adapted from Acharya et al., 2024, Nature communications).

We have deployed the specific property of *FnCas9* and its engineered variant for detecting a wide range of pathogenic Single Nucleotide Variants named it as FELUDA. It utilized the direct *FnCas9* based binding as the readout to detect the nucleobase without acknowledging Cas12/Cas13 based trans-cleavage of reporter detection systems. Semi-quantitative approach of FELUDA can be deployed for various detection platforms and versatile therapeutic applications. We exhibited FELUDA's ability for accurately detecting SNVs, even heterozygosity in case of Sickle Cell Anaemia which could enable early detection and informed decision making in healthcare maintenance. Versatility, Scalable and Rapid testing of FELUDA has facilitated the screening of

large populations, enabled prompt spotting of infected individuals, hence, decreasing transmission rates (Figure 1.7).

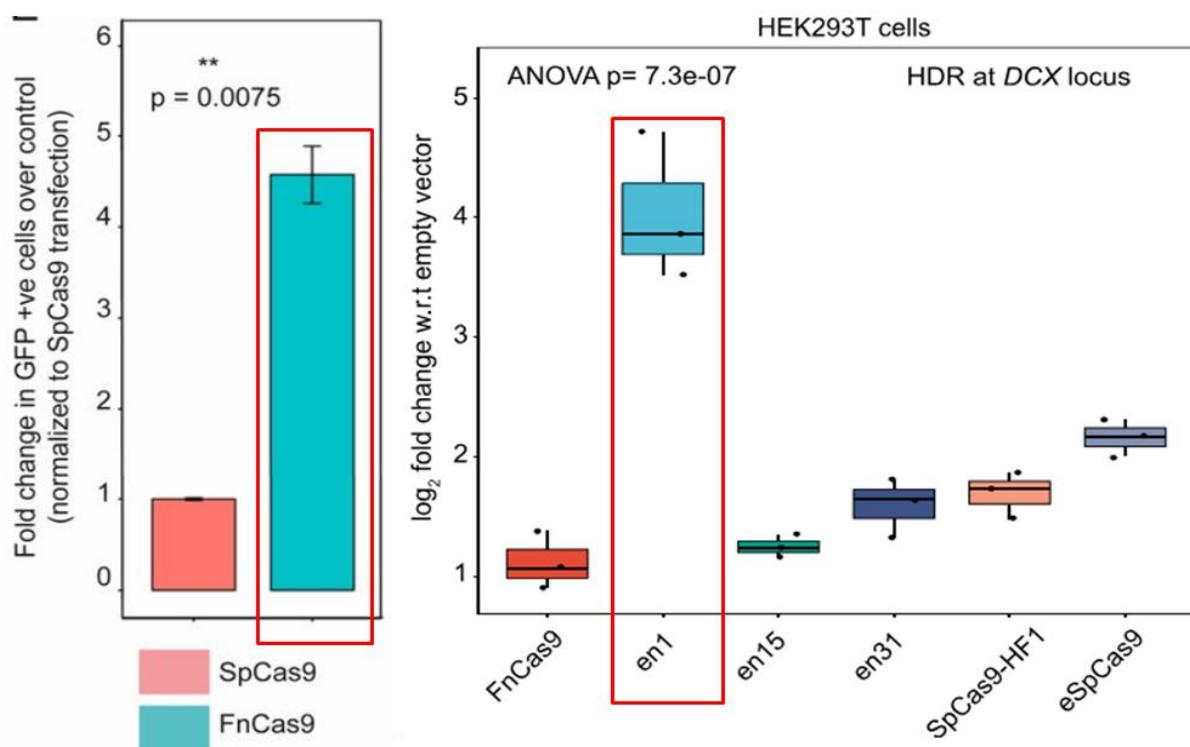
It has shown us better specificity than the Cas13 based assay known as SHERLOCK for SCA detection (Phutela et al., 2022, Springer Protocols; Azhar et al., 2021, BSBE; Kumar et al., 2021, eLife; Acharya et al., 2024, Nature communications).



**Figure 1.7 FnCas9 and its engineered variants can distinguish SNVs via lateral flow assay.** (A) Experimental paradigm showing the chemistry of lateral flow detection platform for distinguishing WT and SNV. (B) Pie chart representing the targetable (yellow) and non-targetable (cyan) percentage of SNVs by FnCas9 and its FnCas9 variants. (C) (Left) Lateral flow detection outcomes of WT and SCD

samples utilizing FnCas9 and engineered FnCas9 FELUDA chemistry, Wild (WT) and Sickle Cell Anemia (SCA) sequences along with the FELUDA specific sgRNAs containing mismatches at their defined positives are represented in red. (Right) Bar graph showing the corresponding TOPSE values written at the bottom of the strips, data is plotted as fold change of SCA sample over its WT type (n=3 independent experiments). Student's unpaired *t*-test p-value is represented for \* <0.05. (Figure adapted from Acharya et al., 2024, Nature communications).

Surprisingly, we have also observed **higher HDR with FnCas9 and one of its engineered variants, *en1FnCas9*** having E1369R amino acid change. This *FnCas9* engineered variant has even shown higher HDR than *SpCas9* high fidelity variants like *SpCas9 HF1* and *eSpCas9* (Figure 1.8).



**Figure 1.8 Bar plot showing the HDR frequencies mediated by FnCas9 and en1FnCas9.**

(Left) HDR represented as fold-change in GFP+ cells over a scrambled sgRNA control (normalized to SpCas9 transfection). Error bars represent SEM (3 independent experiments). Student's *t* test *P* values are shown. (Right) Box plot showing knock-in of a donor template at DCX locus by FnCas9, en1, en15, en31, SpCas9-HF1 and eSpCas9 in HEK293T cells. Data is represented as log<sub>2</sub> fold change w.r.t. empty vectors, and analysed using one-way ANOVA, *p*-value is shown. The middle line within the box represents the median, the box edges represent the interquartile ranges and the whiskers

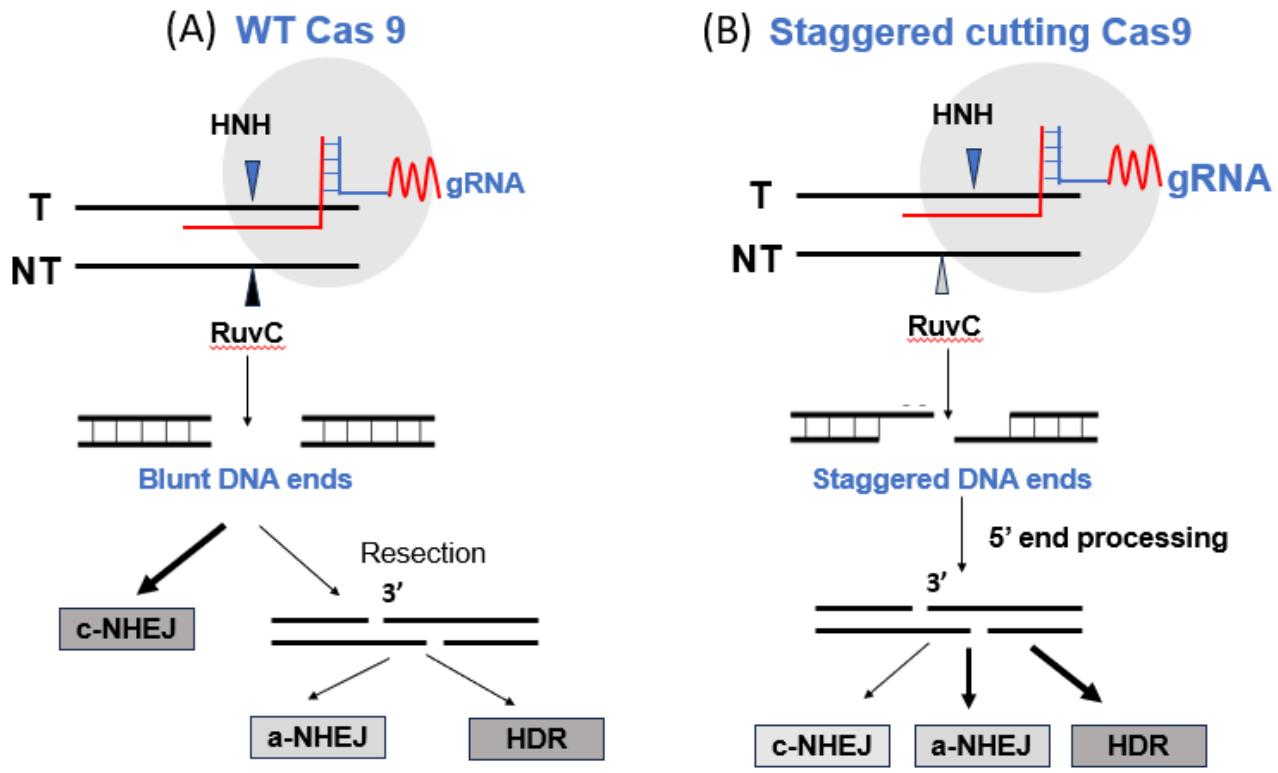
indicate  $\pm 1.5 \times$  interquartile ranges of  $n=3$  independent biological replicates with individual values shown as dots. (Figure adapted from Acharya et al., 2019, PNAS; Acharya et al., 2024, Nature communications).

A prior study also found increase in HDR mediated by Cpf1 protein that identifies a PAM sequence rich in thymine (T), produces **staggered overhangs**, and displays lower tolerance for single or double mismatches in DNA: RNA heteroduplexes, resulting in lower cellular off-targeting. These findings suggest potential similarities between *FnCas9* or its engineered variants and *Cpf1* in their cleavage mechanisms and target preference (Acharya et al., 2019). To explore potential similarities in the cleavage mechanism between *FnCas9* or its engineered variants and *Cpf1*, investigation is needed and this also initiates studies like staggered ends promoting HDR frequencies.

**1.5.8 The structure of DNA breaks influences repair mechanisms:** Earlier, Cas9 has been engineered extensively for enhanced on-target versus off-target DNA cutting specificity and altered protospacer-adjacent motif (PAM) recognition. Some Cas9 variants display slightly different frequencies of DSB repair pathways, but the mechanisms driving these repair biases are not well understood. Particularly, the repair outcomes inside a cell when DNA breaks that are staggered or blunt have not been systematically investigated. To fully harness the potential of CRISPR-Cas9 precise DNA edits, it is crucial to study the nature of the DNA lesions it creates and the repair pathways that the cell's machinery engages to fix these lesions.

It has also been noticed that presence and polarity of overhang structure produced by Cas9 cleavage play significant roles in determining the choice of the DNA repair pathway chosen to fix the DSBs. This has been seen with the nicks generated by the *SpCas9* D10A and N863A nickase variants that activate DNA repair pathways in different manners. Nicks created by the D10A variant favors HR pathways which is in contrast with the N863A. Indicating the role of structure of DNA ends generated and their compatibility with the downstream repair processes (Bothmer et al., 2017) (Figure 1.9).

Blunt ended DSBs created by Cas9 usually lead to insertions or deletions (indels) via non-homologous end joining (NHEJ). Contrarily, the nickase version of Cas9 that produces longer 5' staggered DNA ends favoured pathways like MMEJ and HDR because of the initiation of DNA resection. These studies led to a proposal of at least 6-base pair overhangs as a potential strategy for preferring HR pathways over the classical NHEJ, observed with the *vSpCas9* (Chauhan et al., 2023)



**Figure 1.9 Models depicting the engagement of repair proteins depending upon the ends of the DSBs.** (A) Model depicting the predominant engagement of c-NHEJ for the repair of WT Cas9 induced DSBs. (B) Model depicting the predominant engagement of HDR for the repair of staggered Cas9 induced DSBs

Systematic study of types and lengths of DNA lesions created by CRISPR Cas9 system and its downstream processing via different repair pathways is the crucial area of study to facilitate the therapeutic interventions. Thorough investigations of the mechanistic basis of selection of repair pathways in response to the cleavage induced by Cas9 is indispensable for understanding and optimizing the genome editing strategies.

## 1.6 Aims and objectives

Therefore, this thesis work seems to be centred around **investigating the DNA cleavage induced repair dynamics of FnCas9 and its engineered variants in comparison with SpCas9**. This research is part of the broader effort to fine-tune FnCas9 and its variants CRISPR systems for more precise and efficient

genome editing applications. To strengthen this hypothesis, we formulated the following objectives:

- 1. To investigate the structure of DNA breaks generated by FnCas9/en1 FnCas9 upon DNA cleavage.**
- 2. To comprehend the repair proteins recruitment triggered upon en1FnCas9 and SpCas9 DNA cleavage.**
- 3. To examine the repair outcomes mediated by en1 FnCas9 and SpCas9 at different loci.**

The objectives are presented across four chapters where staggered ended products produced by *FnCas9* and *en1FnCas9* cleavage leads to more efficient dsDNA-break sealing and consequently higher HDR outcomes. These advancements lay the groundwork for the therapeutic genome editing applications.

## 1.7 Bibliography

1. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of bacteriology*, 169(12), pp.5429-5433.
2. Groenen, P.M., Bunschoten, A.E., Soolingen, D.V. and Errtbden, J.D.V., 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular microbiology*, 10(5), pp.1057-1065
3. Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S.J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D. and Musser, J.M., 1999. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerging infectious diseases*, 5(2), p.254.
4. Sorek, R., Kunin, V. and Hugenholtz, P., 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology*, 6(3), pp.181-186.
5. Popkov VA, Zorova LD, Korvigo IO, Silachev DN, Jankauskas SS, Babenko VA, Pevzner IB, Danilina TI, Zorov SD, Plotnikov EY, Zorov DB. Do mitochondria have an immune system? Biochemistry (Moscow). 2016 Oct; 81:1229-36.
6. Makarova, K.S., Wolf, Y.I., Snir, S. and Koonin, E.V., 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology*, 193(21), pp.6039-6056.
7. Marraffini, L.A. and Sontheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in

- staphylococci by targeting DNA. *science*, 322(5909), pp.1843-1845.
- 8. Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K. and Zhang, F., 2015. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Molecular cell*, 60(3), pp.385-397.
  - 9. Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L. and Severinov, K., 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, 353(6299), p. aaf5573.
  - 10. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DB, Kellner MJ, Regev A, Lander ES. RNA targeting with CRISPR–Cas13. *Nature*. 2017 Oct 12;550(7675):280-4.
  - 11. Takeuchi, N., Wolf, Y.I., Makarova, K.S. and Koonin, E.V., 2012. Nature and intensity of selection pressure on CRISPR-associated genes. *Journal of bacteriology*, 194(5), pp.1216-1225.
  - 12. Koonin, E.V. and Makarova, K.S., 2022. Evolutionary plasticity and functional versatility of CRISPR systems. *PLoS Biology*, 20(1), p.e3001481.
  - 13. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P. and Moineau, S., 2020. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18(2), pp.67-83.
  - 14. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E., 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), pp.602-607.
  - 15. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E., 2012. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science*, 337(6096), pp.816-821.
  - 16. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O., 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5), pp.935-949.
  - 17. Gupta, R.M. and Musunuru, K., 2014. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *The Journal of clinical investigation*, 124(10), pp.4154-4161.
  - 18. Singh, A., Chakraborty, D. and Maiti, S., 2016. CRISPR/Cas9: a historical and chemical biology perspective of targeted genome engineering. *Chemical Society Reviews*, 45(24), pp.6666-6684.

19. Chapman, J.R., Taylor, M.R. and Boulton, S.J., 2012. Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*, 47(4), pp.497-510.
20. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), pp.819-823.
21. Symington LS, Gautier J. Double-strand break end resection and repair pathway choice. Annual review of genetics. 2011 Dec 15; 45:247-71.
22. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M., 2013. RNA-guided human genome engineering via Cas9. *Science*, 339(6121), pp.823-826.
23. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J., 2013. RNA-programmed genome editing in human cells. *elife*, 2, p.e00471.
24. Lieber, M.R., Gu, J., Lu, H., Shimazaki, N. and Tsai, A.G., 2010. Nonhomologous DNA end joining (NHEJ) and chromosomal translocations in humans. *Genome stability and human diseases*, pp.279-296.
25. Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lecrivain, A.L., Bzdrenga, J., Koonin, E.V. and Charpentier, E., 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research*, 42(4), pp.2577-2590.
26. Chylinski, K., Makarova, K.S., Charpentier, E. and Koonin, E.V., 2014. Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, 42(10), pp.6091-6105.
27. Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S. and Kim, J.H., 2017. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nature communications*, 8(1), p.14500.
28. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S. and Koonin, E.V., 2015. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520(7546), pp.186-191.
29. Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J. and Xia, N., 2020. Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. *International journal of molecular sciences*, 21(18), p.6461.
30. Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C. and Brouns, S.J., 2017. CRISPR-Cas: adapting to change. *Science*, 356(6333), p. eaal5056.
31. Adhikari, P. and Poudel, M., 2020. CRISPR-Cas9 in agriculture: Approaches, applications, future perspectives, and associated challenges. *Malaysian Journal of Halal Research*, 3(1), pp.6-16.

32. Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J. and Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3), pp.733-740.
33. Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E. and Doudna, J.A., 2016. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, 351(6275), pp.867-871.
34. Peng, Y., Clark, K.J., Campbell, J.M., Panetta, M.R., Guo, Y. and Ekker, S.C., 2014. Making designer mutants in model organisms. *Development*, 141(21), pp.4042-4054.
35. Xue, C. and Greene, E.C., 2021. DNA repair pathway choices in CRISPR-Cas9-mediated genome editing. *Trends in Genetics*, 37(7), pp.639-656.
36. Fell, V.L. and Schild-Poulter, C., 2015. The Ku heterodimer: function in DNA repair and beyond. *Mutation Research/Reviews in Mutation Research*, 763, pp.15-29.
37. Kraglund, B.B., Weterings, E., Hartmann-Petersen, R. and Keijzers, G., 2016. The Ku70/80 ring in non-homologous end-joining: easy to slip on, hard to remove. *Front. Biosci*, 21, pp.514-527.
38. Davis, A.J., Chen, B.P. and Chen, D.J., 2014. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA repair*, 17, pp.21-29.
39. Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K. and Sherwood, R.I., 2018. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*, 563(7733), pp.646-651.
40. Chakrabarti, A.M., Henser-Brownhill, T., Monserrat, J., Poetsch, A.R., Luscombe, N.M. and Scaffidi, P., 2019. Target-specific precision of CRISPR-mediated genome editing. *Molecular cell*, 73(4), pp.699-713.
41. Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M. and Bassett, A.R., 2019. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nature biotechnology*, 37(1), pp.64-72.
42. Brinkman, E.K., Chen, T., de Haas, M., Holland, H.A., Akhtar, W. and van Steensel, B., 2018. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Molecular cell*, 70(5), pp.801-813.
43. Shi, X., Shou, J., Mehryar, M.M., Li, J., Wang, L., Zhang, M., Huang, H., Sun, X. and Wu, Q., 2019. Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. *Cell discovery*, 5(1), p.53.
44. Clouaire, T. and Legube, G., 2019. A snapshot on the cis chromatin response to DNA double-strand breaks. *Trends in Genetics*, 35(5), pp.330-345.

45. Isaac, R.S., Jiang, F., Doudna, J.A., Lim, W.A., Narlikar, G.J. and Almeida, R., 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife*, 5, p.e13450.
46. Schep, R., Brinkman, E.K., Leemans, C., Vergara, X., van der Weide, R.H., Morris, B., van Schaik, T., Manzo, S.G., Peric-Hupkes, D., van den Berg, J. and Beijersbergen, R.L., 2021. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Molecular cell*, 81(10), pp.2216-2230.
47. Doudna, J.A. and Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213), p.1258096.
48. Sung, P. and Klein, H., 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nature reviews Molecular cell biology*, 7(10), pp.739-750.
49. Bhargava, R., Onyango, D.O. and Stark, J.M., 2016. Regulation of single-strand annealing and its role in genome maintenance. *Trends in Genetics*, 32(9), pp.566-575.
50. Yeh, C.D., Richardson, C.D. and Corn, J.E., 2019. Advances in genome editing through control of DNA repair pathways. *Nature cell biology*, 21(12), pp.1468-1478.
51. Cox, D.B.T., Platt, R.J. and Zhang, F., 2015. Therapeutic genome editing: prospects and challenges. *Nature medicine*, 21(2), pp.121-131.
52. WareJoncas, Z., Campbell, J.M., Martínez-Gálvez, G., Gendron, W.A., Barry, M.A., Harris, P.C., Sussman, C.R. and Ekker, S.C., 2018. Precision gene editing technology and applications in nephrology. *Nature Reviews Nephrology*, 14(11), pp.663-677.
53. Doudna JA. The promise and challenge of therapeutic genome editing. *Nature*. 2020 Feb 13;578(7794):229-36.
54. Barrangou, R. and Doudna, J.A., 2016. Applications of CRISPR technologies in research and beyond. *Nature biotechnology*, 34(9), pp.933-941.
55. Garcia-Robledo, J.E., Barrera, M.C. and Tobón, G.J., 2020. CRISPR/Cas: from adaptive immune system in prokaryotes to therapeutic weapon against immune-related diseases: CRISPR/Cas9 offers a simple and inexpensive method for disease modelling, genetic screening, and potentially for disease therapy. *International Reviews of Immunology*, 39(1), pp.11-20.
56. Vartak, S.V. and Raghavan, S.C., 2015. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. *The FEBS journal*, 282(22), pp.4289-4294.
57. Singh, P., Schimenti, J.C. and Bolcun-Filas, E., 2015. A mouse geneticist's practical guide to CRISPR applications. *Genetics*, 199(1), pp.1-15.
58. Zhang, Y., Zhang, Z. and Ge, W., 2018. An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing. *Journal of*

*Biological Chemistry*, 293(17), pp.6611-6622.

59. Ma, Y., Chen, W., Zhang, X., Yu, L., Dong, W., Pan, S., Gao, S., Huang, X. and Zhang, L., 2016. Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. *RNA biology*, 13(7), pp.605-612.
60. Hu, Z., Shi, Z., Guo, X., Jiang, B., Wang, G., Luo, D., Chen, Y. and Zhu, Y.S., 2018. Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR–Cas9 and ssODN in human cancer cells. *Cell & bioscience*, 8, pp.1-15.
61. Li, G., Zhang, X., Zhong, C., Mo, J., Quan, R., Yang, J., Liu, D., Li, Z., Yang, H. and Wu, Z., 2017. Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Scientific Reports*, 7(1), p.8943.
62. Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R. and Ploegh, H.L., 2015. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nature biotechnology*, 33(5), pp.538-542.
63. Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kühn, R., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nature biotechnology*, 33(5), pp.543-548.
64. Li, G., Liu, D., Zhang, X., Quan, R., Zhong, C., Mo, J., Huang, Y., Wang, H., Ruan, X., Xu, Z. and Zheng, E., 2018. Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *The international journal of biochemistry & cell biology*, 99, pp.154-160.
65. Robert, F., Barbeau, M., Éthier, S., Dostie, J. and Pelletier, J., 2015. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome medicine*, 7, pp.1-11.
66. Frank, K.M., Sekiguchi, J.M., Seidl, K.J., Swat, W., Rathbun, G.A., Cheng, H.L., Davidson, L., Kangaloo, L. and Alt, F.W., 1998. Late embryonic lethality and impaired V (D) J recombination in mice lacking DNA ligase IV. *Nature*, 396(6707), pp.173-177.
67. Pinder, J., Salsman, J. and Dellaire, G., 2015. Nuclear domain ‘knock-in’ screen for the evaluation and identification of small molecule enhancers of CRISPR-based genome editing. *Nucleic acids research*, 43(19), pp.9379-9392.
68. Kurihara, T., Kouyama-Suzuki, E., Satoga, M., Li, X., Badawi, M., Baig, D.N., Yanagawa, T., Uemura, T., Mori, T. and Tabuchi, K., 2020. DNA repair protein RAD51 enhances the CRISPR/Cas9-mediated knock-in efficiency in brain neurons. *Biochemical and biophysical research communications*, 524(3), pp.621-628.
69. Paulsen, B.S., Mandal, P.K., Frock, R.L., Boyraz, B., Yadav, R., Upadhyayula, S., Gutierrez-Martinez, P., Ebina, W., Fasth, A., Kirchhausen, T. and Talkowski, M.E., 2017. Ectopic expression

- of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. *Nature biomedical engineering*, 1(11), pp.878-888.
70. Charpentier, M., Khedher, A.H.Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., Geny, S. and De Cian, A., 2018. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nature communications*, 9(1), p.1133.
71. Tran, N.T., Bashir, S., Li, X., Rossius, J., Chu, V.T., Rajewsky, K. and Kühn, R., 2019. Enhancement of precise gene editing by the association of Cas9 with homologous recombination factors. *Frontiers in genetics*, 10, p.365.
72. Reuven, N., Adler, J., Broennimann, K., Myers, N. and Shaul, Y., 2019. Recruitment of DNA repair MRN complex by intrinsically disordered protein domain fused to Cas9 improves efficiency of CRISPR-mediated genome editing. *Biomolecules*, 9(10), p.584.
73. Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S. and Qi, L.S., 2015. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell stem cell*, 16(2), pp.142-147.
74. Jayavaradhan, R., Pillis, D.M., Goodman, M., Zhang, F., Zhang, Y., Andreassen, P.R. and Malik, P., 2019. CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nature communications*, 10(1), p.2866.
75. Nambiar, T.S., Billon, P., Diedenhofen, G., Hayward, S.B., Taglialatela, A., Cai, K., Huang, J.W., Leuzzi, G., Cuella-Martin, R., Palacios, A. and Gupta, A., 2019. Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nature Communications*, 10(1), p.3395.
76. Bothmer, A., Phadke, T., Barrera, L.A., Margulies, C.M., Lee, C.S., Buquicchio, F., Moss, S., Abdulkerim, H.S., Selleck, W., Jayaram, H. and Myer, V.E., 2017. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nature communications*, 8(1), p.13905.
77. Lin, S., Staahl, B.T., Alla, R.K. and Doudna, J.A., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *elife*, 3, p.e04766.
78. Yang, D., Scavuzzo, M.A., Chmielowiec, J., Sharp, R., Bajic, A. and Borowiak, M., 2016. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Scientific reports*, 6(1), p.21264.
79. Lomova, A., Clark, D.N., Campo-Fernandez, B., Flores-Bjurström, C., Kaufman, M.L., Fitz-Gibbon, S., Wang, X., Miyahira, E.Y., Brown, D., DeWitt, M.A. and Corn, J.E., 2019. Improving gene editing outcomes in human hematopoietic stem and progenitor cells by temporal control of DNA repair. *Stem cells*, 37(2), pp.284-294.

80. Wienert, B., Nguyen, D.N., Guenther, A., Feng, S.J., Locke, M.N., Wyman, S.K., Shin, J., Kazane, K.R., Gregory, G.L., Carter, M.A. and Wright, F., 2020. Timed inhibition of CDC7 increases CRISPR-Cas9 mediated templated repair. *Nature communications*, 11(1), p.2109.
81. Devkota, S., 2018. The road less travelled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB reports*, 51(9), p.437.
82. Carlson-Steumer, J., Abdeen, A.A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M. and Saha, K., 2017. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nature communications*, 8(1), p.1711.
83. Ma, M., Zhuang, F., Hu, X., Wang, B., Wen, X.Z., Ji, J.F. and Xi, J.J., 2017. Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Research*, 27(4), pp.578-581.
84. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. and Corn, J.E., 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology*, 34(3), pp.339-344.
85. Song, F. and Stieger, K., 2017. Optimizing the DNA donor template for homology-directed repair of double-strand breaks. *Molecular Therapy-Nucleic Acids*, 7, pp.53-60.
86. Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F. and Nishimasu, H., 2016. Structure and engineering of Francisella novicida Cas9. *Cell*, 164(5), pp.950-961.
87. Acharya, S., Mishra, A., Paul, D., Ansari, A.H., Azhar, M., Kumar, M., Rauthan, R., Sharma, N., Aich, M., Sinha, D. and Sharma, S., 2019. Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proceedings of the National Academy of Sciences*, 116(42), pp.20959-20968.
88. Chauhan, V.P., Sharp, P.A. and Langer, R., 2023. Altered DNA repair pathway engagement by engineered CRISPR-Cas9 nucleases. *Proceedings of the National Academy of Sciences*, 120(11), p.e2300605120.
89. Matsumoto, D., Matsugi, E., Kishi, K., Inoue, Y., Nigorikawa, K. and Nomura, W., 2024. SpCas9-HF1 enhances accuracy of cell cycle-dependent genome editing by increasing HDR efficiency, and by reducing off-target effects and indel rates. *Molecular Therapy-Nucleic Acids*, 35(1).

## **Chapter 2: To investigate the structure of DNA breaks generated by FnCas9/en1 FnCas9 upon DNA cleavage**

*An overview of how FnCas9 and its engineered variant create 2-3 base staggered DNA ends during in-vitro and in-cellulo conditions and processing of its ends are different from the blunt ended DNA.*

## 2.1 Background

Resolving the choice of DNA repair pathways relies on phases of cell cycle and properties of DNA end resection which determine the accuracy and efficiency of repair outcomes. The time taken by cells to detect and repair Cas9-induced DSBs may vary from naturally occurring DSBs. During genome editing Cas9 can remain bound to the DNA for a period of time which impacts the repair kinetics and choice of DNA repair machinery (Sternberg et al., 2014). Certainly, DSBs are primarily the most harmful lesions among others like SSBs, mismatches, modified bases, and abasic sites possessed by Ionizing radiations (Zhao et al., 2020). Despite c-NHEJ being the prominent repair pathway for DSBs induced by ionizing radiation, altered repair kinetics of Cas9-induced DSBs may favour pathways like MMEJ over c-NHEJ. DSBs generated by Cas9 are majorly blunt ends which can be directly ligated through c-NHEJ. Less frequent staggered ends (1-2 bases at 5' ends) may require further processing before repair that impacts the kinetics and choice of the repair pathways. Staggered ends are proposed to slow down the repair kinetics via c-NHEJ and this time delay may be considerable for cells to commence short-range end resection or long-range end resection to promote the MMEJ and HDR pathways respectively (Shou et al., 2018).

Occasionally induced large deletions, extending several kilobases by Cas9 could possess serious consequences in clinical settings. Inquiring mechanisms supporting these deletions, repeatedly linked to p53-dependent mechanisms, is pivotal to alleviate potential side effects in genome editing (Kosicki et al., 2018; Adikusuma et al., 2018). Comprehending these complexities in DSBs detection times, repair pathways and unintended outcomes is central for proceeding CRISPR-Cas-based genome editing toward precise and safer clinical applications. Research efforts focused on clarifying these mechanisms will be vital to enhance the reliability and safety of genome editing techniques for therapeutic purposes (Korablev et al., 2020; Cullot et al., 2019).

Previous studies have shown higher HDR mediated by FnCas9 and its engineered variants en1FnCas9 (Acharya et al., 2019; Acharya et al., 2024). Speculating the mutations that have been done for creating a potential Cas9 having precise and safe on-target genome editing, it's important to observe its DNA cleavage mode as a reason for increased HDR.

## 2.2 Methods and Materials

### 2.2.1 Plasmid Construction

PX408 (Addgene 68705) was used as a template for PCR amplification of the gene encoding the full length *Francisella novicida Cas9* (*FnCas9*) nuclease residues from 1–1629 bp as well as its modified variations. Following that, these genes were cloned using restriction enzyme-based cloning for the pET28-His-10-Smt3 vector (a gracious gift from Prof. Stewart Shuman and Dr. K.M. Sinha) and ligation-independent cloning (LIC) for the pET-His6-GFP-TEV-LIC vector (Addgene 29663), respectively. With minor changes to the manufacturer's procedure, QuickChange II site directed mutagenesis kit (Agilent) was used to create catalytically inactive FnCas9 double mutants on pET His6-FnGFP-TEV-LIC plasmid backbone.

### 2.2.2 Protein purification

As stated earlier, the proteins employed in this investigation were purified. To put it briefly, *Escherichia coli Rosetta2 (DE3)* (Novagen) was used for creating plasmids encoding distinct Cas9 proteins. Until the OD600 reached 0.6, the *Rosetta2 (DE3)* cells were grown in LB media supplemented with 50 mg/l kanamycin at 37°C. The addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) was used to stimulate protein expression. After an overnight cultivation at 18°C, the cells were separated by centrifugation. Re-suspended *E. Coli* cells were lysed by sonication and centrifugation in lysis buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol), supplemented with 1X protease inhibitor cocktail (PIC, Roche), and 100 µg/ml lysozyme. The lysate was affinity-purified by Ni-NTA beads from Roche, and the eluted protein was further purified using chromatography based on size-exclusion on a HiLoad Superdex 200 (16/60 column) from GE Healthcare in 150 mM KCl, 20 mM HEPES pH 7.5, 1 mM DTT and 10% glycerol. Thermo Fisher Scientific's Pierce BCA protein assay kit was used to quantify the concentration of purified proteins. Up until their next usage, the purified proteins were kept in storage at -80°C. For 6XHis-MBP-dSpCas9, dFnCas9, and its engineered variants the 6XHis-MBP was eliminated by soaking the affinity-bound protein in cleavage buffer (50 mM Tris-Cl, pH8, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT) for a whole night in order to incubate the protein with PreScission Protease. On a HiLoad Superdex 200 16/60 column (GE Healthcare) with 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, and 1 mM DTT, the cleaved Cas9 protein was isolated from the fusion tag.

### **2.2.3 *In vitro* transcription**

Using the Mega Script T7 kit (Thermo Fisher Scientific) and T7 promoters with templates as substrates, sgRNAs were transcribed in vitro. IVT reactions were purified using a NucAway spin column (Thermo Fisher Scientific) after being incubated for a full night at 37°C. Until they were needed again, IVT sgRNAs were kept at -20°C.

### **2.2.4 *In vitro* Cleavage (IVC) assay**

The Cas9-sgRNA as RNP complex was reconstituted in reaction buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM DTT, 10% glycerol, and 10 mM MgCl<sub>2</sub>) at 25°C for 10 mins. A reconstituted RNP complex (50 nM) and the necessary quantity of DNA substrates (5 nM) were incubated at 37°C in a reaction buffer. The corresponding figure legends provided information on the various experiments. Samples were treated for 15 mins at 55°C with 1 ul of 20 mg/ml Proteinase K in a 15 ul reaction. The reaction products were measured using ImageJ, visualised using a Syngene UV transilluminator, and resolved on a 1% agarose gel stained with EtBr. The IVC assay's matching substrates were noted in the legend of figure, and the reactions were terminated by putting 25 mM EDTA as a final concentration.

### **2.2.5 Cloning of IVC products**

500 ng of PCR product for each locus was added to 500 nM of Cas9-sgRNA complex, that had been reconstituted in reaction buffer for a 10-mins period at 25°C. Products which were cleaved in vitro were gel purified using a Qiagen QI quick Gel extraction kit. After gel extraction, the products were cloned using the sticky-end cloning technique of the Clone JET PCR Cloning kit (Thermo Fisher Scientific) in the pJET 1.2/blunt cloning vector. Chemically competent E. coli cells were transformed using the recombinant pJET 1.2/blunt cloning vector. In LB medium containing 100 ug/ml ampicillin, transformed colonies were inoculated, and the QIAprep spin Miniprep kit (Qiagen) was used to purify the colonies. Sanger sequencing and PCR screening were used to test the purified plasmids (AgriGenome Labs).

### **2.2.6 Purification of the Cas9 cut linear products**

Reconstituted 500 nM of Cas9-sgRNA complex was incubated in a reaction buffer at 25°C for 10 mins, using 500 ng of plasmid for every locus. Using a QI quick Gel extraction kit (Qiagen), in vitro cleaved linearized plasmids were gel eluted. Before being transformed or transfected into cells, an equal number

of these products were examined on an agarose gel and quantified using the Qubit® dsDNA BR Assay, per the manufacturer's

### **2.2.7 Uptake of exogenous templates by bacterial and mammalian cells**

After quantification of the Cas9 cleaved products by Qubit, equal molecules of these blunt and staggered ended exogenous templates have been transformed into chemically competent DH5 $\alpha$  *E.coli* cells. Colonies that have survived after efficient sealing of ampicillin selection markers were counted manually and plotted after 20-24 hours of the incubation in the bacterial incubators. Similarly, equal molecules of Cas9 cut products have been transfected in mammalian cells like HEK293T cells via lipofectamine p3k reagent. The repair accuracy was calculated by quantifying the GFP positive cells after 2 days of the incubation using FACS sorter from BD melody (BD Biosciences-US).

### **2.2.8 Guide RNA cloning**

To produce the crRNA in the cells we have cloned the spacer sequence (20 nucleotides) of targets like *EMX*, *HBB*, *FASN*, *LMNA*, *OCT4* under the U6 promoter having a bbs1 crRNA cloning site in the construct expressing GFP under strong promoter. en1FnCas9 and SpCas9 Tracr scaffold constructs have been sequence confirmed and their GFP signal has been confirmed via transfection in HEK293T cells.

### **2.2.9 Sanger Sequencing**

The sequencing reaction was carried out using big dye Terminator v3.1 cycle sequencing kit (ABI, 4337454) in 10  $\mu$ l volume. (containing 0.5  $\mu$ l purified DNA, 0.8  $\mu$ l sequencing reaction mix, 2  $\mu$ l 5X dilution buffer, and 0.6  $\mu$ l forward/ reverse primer) with the following cycling conditions - 3 mins at 95°C, 40 cycles of (10 sec at 95°C, 10 sec at 55°C, 4 min at 60°C) and 10 mins at 4°C. Subsequently, the PCR product was purified by mixing with 12  $\mu$ l of 125 mM EDTA (pH 8.0) and incubating at RT for 5 mins. 50  $\mu$ l of absolute ethanol and 2  $\mu$ l of 3 M NaOAc (pH 4.8) was then added, incubated at RT for 10 mins and centrifuged at 3800 rpm for 30 mins, followed by invert spin at <300 rpm to discard the supernatant. The pellet was washed twice with 100  $\mu$ l of 70% ethanol at 4000 rpm for 15 mins, and the supernatant was discarded by invert spin. The pellet was air-dried, dissolved in 12  $\mu$ l of Hi-Di formamide (Thermo fisher, 4311320), denatured at 95°C for 5 mins followed by a snap chill, and linked to ABI 3130xl sequencer. Base-calling was carried out using sequencing analysis software (v5.3.1) (ABI, US), and the sequence was analysed using Chromas v2.6.5 (Technelysium, Australia)

## **2.2.10 Data Analysis**

Determination of indel frequency from sequencing data was performed using CRISPResso2 v2.0.29 (52) with the following parameters ‘--ignore\_substitution -- min\_paired\_end\_reads\_overlap 10 -- max\_paired\_end\_reads\_overlap of 500. We have detected by sanger sequencing that the cleavage positions for en1 FnCas9 on the non-target strand varied from 3-8bp upstream of the PAM, so, the quantification window for indel detection was set 3 to 8 bp upstream of PAM.

## **2.2.11 Cell culture and electroporation**

HEK 293T cells were grown in DMEM media supplemented with high glucose (Invitrogen), 2 mM GlutaMax, 10% FBS (Invitrogen), 1X antibiotic and antimycotic (Invitrogen) at 37°C in 5% CO<sub>2</sub>. Electroporation of mammalian cells were performed using Neon™ Transfection System 10 μL Kit following the manufacturer's protocol. For each electroporation reaction, total 1 ug Cas9-gRNA vector plasmid was mixed in Resuspension buffer R. 2 x 10<sup>5</sup> cells were resuspended in 10 μL of Resuspension Buffer R containing DNA and electroporation was performed using Neon® Transfection System 10 ul Kit (ThermoFisher Scientific Cat. No. MPK10096) with single pulses at 950 V, 30 milliseconds pulse width. The electroporated cells were transferred immediately to a 12 well plate containing 1 ml of pre-warmed culture medium and incubated at 37°C and 5% CO<sub>2</sub>. After 24 hours cells were washed and re-incubated with fresh culture medium. 72 hours of post electroporation GFP positive cells per sample were sorted using BD FACSMelody Cell Sorter (BD Biosciences-US) and gDNA was isolated from the sorted cells using Lucigen buffer for NGS sequencing.

## **2.2.12 Amplicon Sequencing**

HEK293T cells on six well dishes were transfected with total 1 ug Cas9-gRNA vector plasmid. 72 hrs of post-transfection GFP-positive cells were FACS sorted (BD FACSMelody Cell Sorter) and gDNA was isolated (Lucigen Quick Extract Extraction solution). Amplicons were prepared with site-specific primers and were indexed with Illumina TruSeq i5 and i7 index adapters. Indexed amplicons were purified with AmPure XP beads. Concentrations were determined with Thermo HS-assay kit (Thermo scientific) in a Qubit instrument and amplicons were pooled to make a sequencing library. Pooled sequencing library was sequenced on an Illumina Novaseq 6000 instrument with 2x150 bp chemistry at required depth.

Reads were demultiplexed on the instrument. Sequencing reads were quality-checked with FastQC

v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.14 (<http://dx.doi.org/10.1093/bioinformatics/btw354>). Poor-quality bases with Phred score less than Q30 and TruSeq adapter sequences were removed using cutadapt v2.8 (<http://dx.doi.org/10.14806/ej.17.1.200>) for paired-end reads. Trimmed reads with at least 45 bp length were retained. To check for indels, cleaned reads were analyzed with CRISPResso2 v2.2.14 (<https://anaconda.org/bioconda/crispresso2>) with default options, except “-qwc” flag was used to restrict the quantification window to the gRNA region and “--ingore\_substitutions”. CRISPR-DAV v2.3.4 (<https://doi.org/10.1093/bioinformatics/btx518>) was used for indel and HDR quantification with default options. An in-house Python script was used to combine editing outcomes from different samples into a CSV file. Data analysis was done in R v4.3.2 and plotted using ggpubr v0.6.0.

Quantification of insertion types: For each target site, cleaned paired-end reads were analysed for indels with CRISPResso2 with above mentioned options and resulting “Alleles\_frequency\_table\_around\_sgRNA\_[sgRNA\_seq].txt” file was used for downstream analysis. A custom in-house Python script was used to quantify what base was inserted at the cleavage site during DNA-repair and the associated frequency. Briefly, it compares the wild-type allele against edited alleles with insertion and looks for inserted bases by alignment. Successively it outputs all the different types of insertions and their frequency.

Quantification of sequence-specific insertion types: For each target site, cleaned paired-end reads were analysed for indels with CRISPResso2 with above mentioned options and resulting “Alleles\_frequency\_table\_around\_sgRNA\_[sgRNA\_seq].txt” file was used for downstream analysis. A custom in-house Python script was used to quantify what base was inserted at the cleavage site during DNA-repair and the associated frequency. First, it compares the wild-type allele against edited alleles harbouring insertion and looks for inserted bases. If the inserted base(s) is exactly the same as the upstream sequence it was considered as “influenced insertion”. To calculate the percentage of such insertions, reads supporting each of those insertions were divided by the total of all reads that contain insertion and were multiplied by hundred.

Formally, list of alleles containing insertion(s):

**A:** = [a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, ..., a<sub>N</sub>],

$\mathbf{A} \subseteq \mathbf{U}$  where  $\mathbf{U}$  is the list of edited alleles containing insertion, deletion or both (from Alleles\_frequency\_table\_around\_sgRNA\_[sgRNA\_seq].txt file).

list of alleles containing influenced insertion(s):  $\mathbf{B} = [b \mid "a" \text{ containing influenced insertion(s), } a \in \mathbf{A}]$   
thus,  $\mathbf{B} \subseteq \mathbf{A}$  and,  $\mathbf{B} = [b_1, b_2, b_3, \dots, b_M], M \leq N$

the percentage of allele “ $b_j$ ” among all insertion-containing alleles:  $\text{Perc}(b_j) = (\text{reads}(b_j) / \sum_{a \in A} \text{reads}(a)) \times 100$ .

Finally, the insertion containing allele, base(s) incorporated and their percentage were plotted.

### 2.2.13 Statistical Analysis

Data are represented as either mean  $\pm$  SD or mean  $\pm$  SEM and indicated in the respective figure legends. Statistical analysis was done using Student’s unpaired t-test. Calculated p-values are represented as follows: \* $P \leq 0.05$ , \*\* $P \leq 0.01$ , \*\*\* $P \leq 0.001$ , and \*\*\*\* $P \leq 0.0001$ .

## 2.3 Results

### 2.3.1 en1FnCas9 generates sticky DNA ends during *in-vitro* and *in-cellulo* conditions

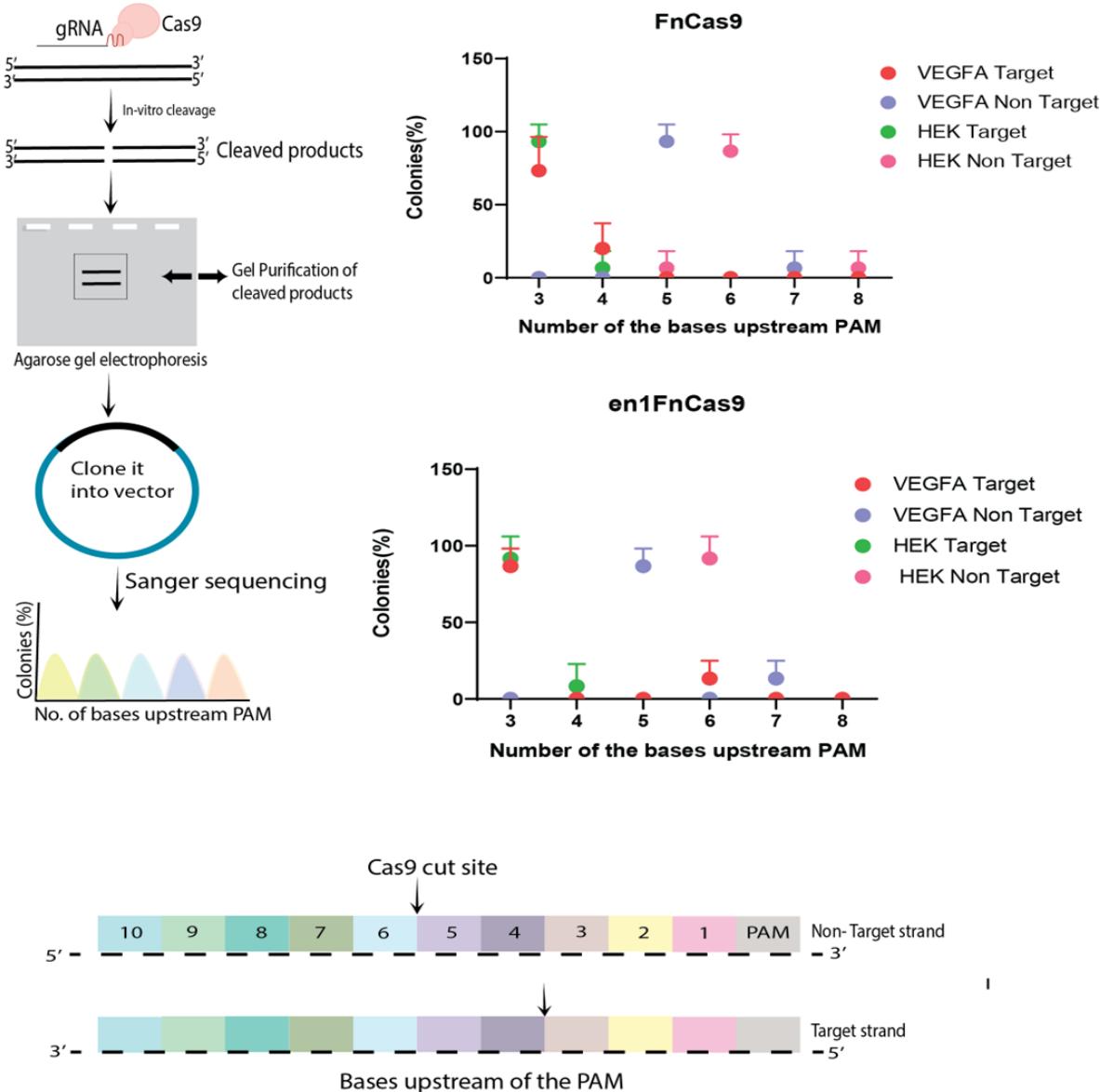
First and foremost, we conducted the research on the cleavage properties of *FnCas9* and its engineered variant *en1FnCas9* in comparison with the canonical *SpCas9* (that creates majority of blunt DNA ends), targeting loci like *VEGFA3*, and *HEK site4*. The DNA products generated by *in-vitro* cleavage reaction were cloned in the destination vector for performing bidirectional sanger sequencing. The observation of flexible staggered non-target strand cleavage of *FnCas9* and its engineered variants was indeed intriguing, especially considering its distinction from other naturally occurring Cas9 proteins.

We have noticed that in nearly all the clones, the target stand of the DNA was cleaved at 3 bp upstream of the PAM by *en1FnCas9*, similar to the canonical *SpCas9*. However, there were distinct non-target strand cleavage positions by *en1FnCas9* starting from (3 to 8 bp away from the PAM), depending upon the sequence, where *SpCas9* cuts 3 bases upstream (Figure 2.1).

Similar staggered DNA cleavage has already been reported by *Cpf1* (*Cas12*), as it recognises T rich

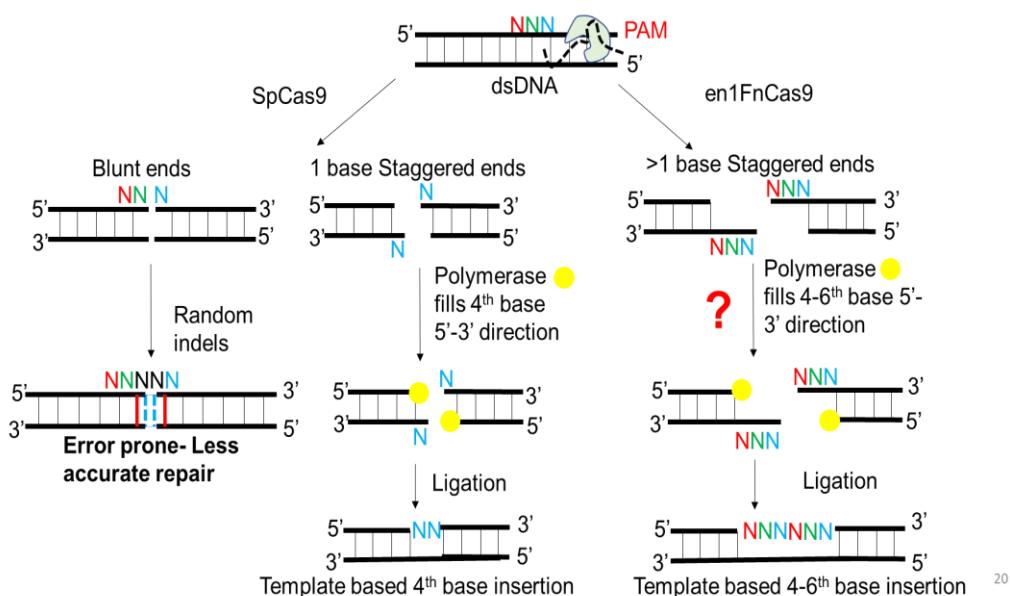
PAM and is also less tolerant to single or double mismatches. Additionally, *Cpf1* has also been associated with less off-targeting in mammalian cells and higher HDR (Zetsche et al., 2020; Mateos et al., 2017). These findings suggested unique mechanisms for staggered ended cleavage by en1FnCas9, that could have significant implications for genome editing applications.

**For our better understanding of specificity and efficiency of *en1FnCas9*, we were curious to know its DNA cleavage pattern in cells as well.** Investigating the cleavage properties and off-target effects of distinct CRISPR Cas systems is vital for improving their specificity and efficiency.



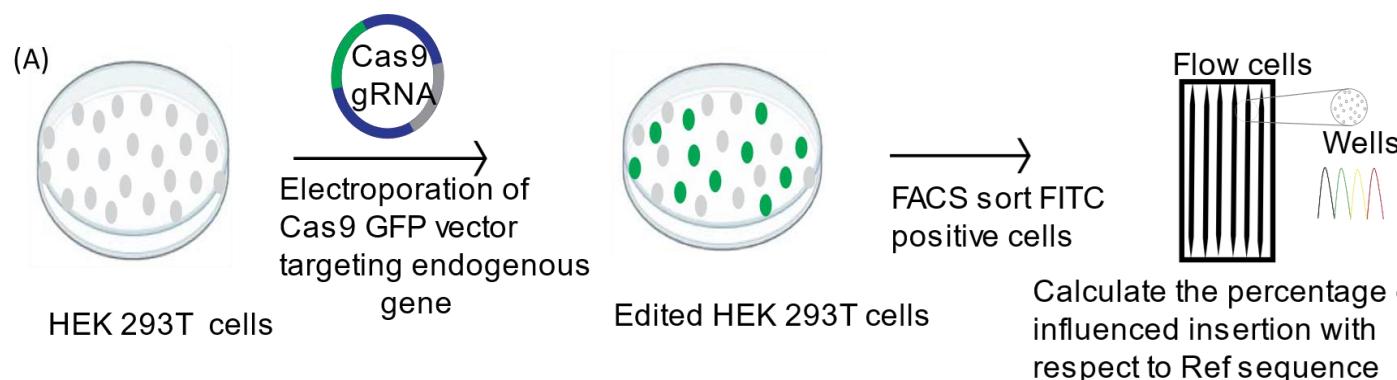
**Figure 2.1 FnCas9 and its engineered variants create 2-3 base staggered DNA ends under *in-vitro* conditions.** Experimental paradigm. The DNA substrates containing the targeted regions are cleaved by *SpCas9* and *en1FnCas9* in *in-vitro* conditions and gel extracted and cloned into vectors for sanger sequencing (Left Panel). Cleavage positions on target and nontarget strand by *FnCas9* and *en1 FnCas9* determined using Sanger sequencing for different targets (*VEGFA* and *HEK*) indicated. The y axis represents the percentage of sequenced clones showing cleavage at a given nucleotide. The x axis represents the position of the base away from PAM in the sgRNA (Right panel). Schematic illustration of *FnCas9* and *en1 FnCas9* cleavage site on 3 base and 5-6 base upstream of PAM on Target and Non-Target strand respectively (Bottom Panel).

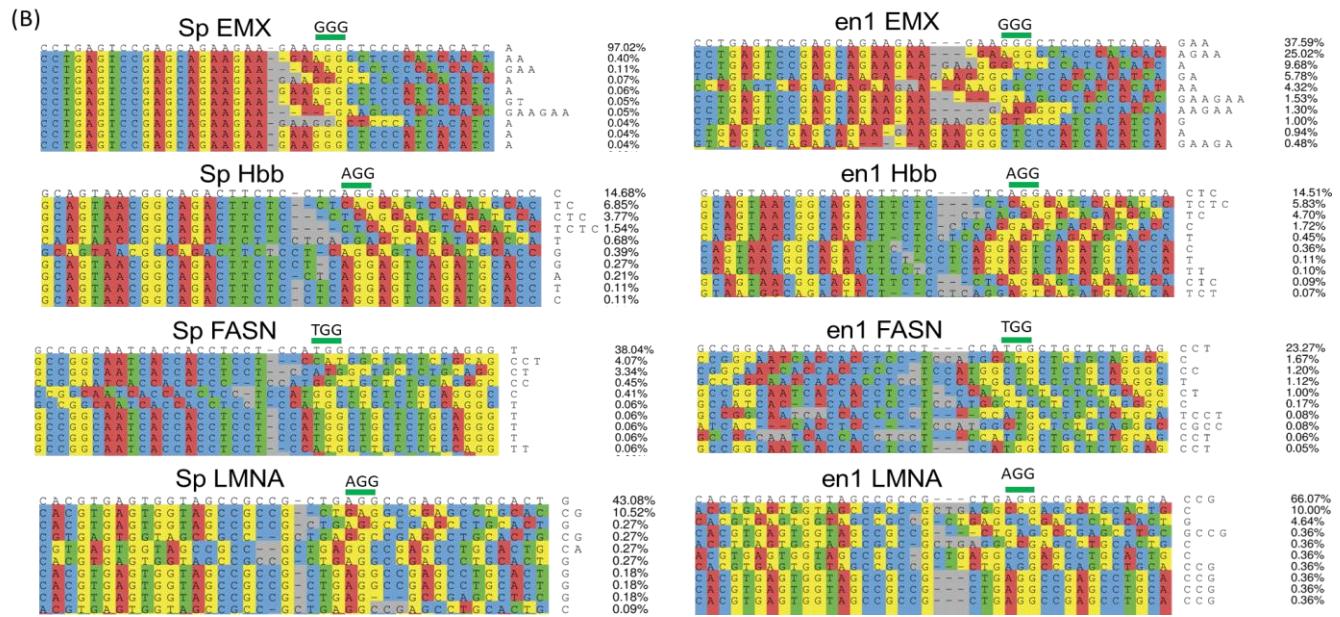
Earlier, it is known that Cas9 majorly induces blunt-ended DSBs which are repaired in a template - independent manner. Recent simulations and evidence on CRISPR-Cas9 systems suggests that Cas9 may also generate staggered DSBs (Shi et al., 2019). This is specifically by HNH and RuvC domains of Cas9, which are capable of cleaving the non-target DNA (ntDNA) and the target DNA (tDNA) between nucleotides 4|5 and 3|4 respectively, upstream of the PAM (Protospacer Adjacent Motif) sequence. Few reports propose polymerase- based fill-in of the resulting 5' overhangs generated by Cas9 cutting at position 4 of the target DNA that produces sgRNA specific insertions by Pol  $\mu$  (DNA polymerase mu) and Pol  $\lambda$  (DNA polymerase lambda) (Gilser et al., 2019). The bases that incorporate during the repair are exactly the same as the preceding base, hence dictates the length of the staggered DNA ends. Therefore, this motivated us to analyze the insertion patterns and sizes induced by *en1 FnCas9* in cells that could depict the length of staggered DNA ends (Figure 2.2).



**Figure 2.2 Experimental paradigm of template dependent polymerase filling in case of Cas9 induced DNA repair.** (Left Panel) SpCas9 creates a majority of blunt DNA ends that leads to random insertions (error prone) and less frequent 1 base staggered DNA ends that leads to single base insertion (4th base upstream) by polymerase filling (yellow) followed by ligation, which is template dependent and accurate. (Right Panel) en1FnCas9 creates 2-3 base staggered DNA ends, have the potential to fill-in by the polymerase and ligate by ligase which leads to bi or trinucleotide (4-5 or 4-6 bases) insertions upstream of PAM which would be accurate.

For analysing the insertional patterns, we electroporated the SpCas9 and en1FnCas9 nuclease plasmids tagged with GFP and targeting endogenous genes like *Fatty acid synthase (FASN)*, *Lamin (LMNA)*, *Haemoglobin (HBB)* and *Empty Spiracles Homeobox 1 (EMX)*. After 24 hours of electroporation, to avoid the electroporation bias in the cells, only GFP positive cells were sorted by FACS that had the Cas9 RNP expression for editing. We then proceeded with the amplicon sequencing to evaluate the percentage of the **influenced insertions**, in which the inserted base(s) is exactly the same as the upstream sequence near the cut site and then we plotted the percentage of the influenced insertion in both the *SpCas9* and *en1Cas9* targeting.





**Figure 2.3 Workflow for investigating en1FnCas9 mediated influenced insertions.** (A) HEK293T cells were electroporated with Cas9 plasmids targeting endogenous genes, GFP positive cells are sorted and percentage of influenced insertion is calculated via deep sequencing. (B) Allele plot showing the percentage of the inserted base(s) upstream of the PAM which is underlined as green colour with its sequences, targeting *EMX*, *HBB*, *FASN* and *LMNA* locus by SpCas9 and en1FnCas9. (Figure generated by Prosad Das).

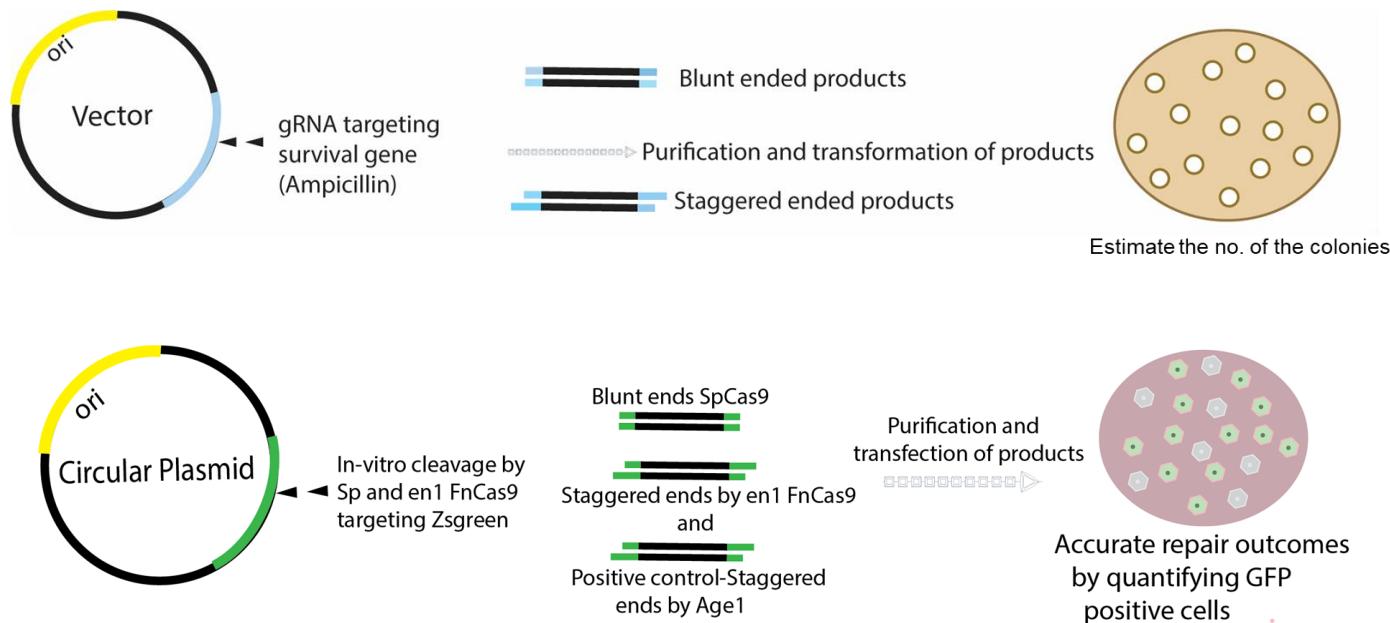
We found frequency of 3 bases insertion is more in case of *en1FnCas9* mediated editing than *SpCas9* in loci like *HBB*, *EMX*, *FASN* and *LMNA*, where we have also observed single base insertion similar to other few reports with *SpCas9*. These 3 bases were exactly similar to the upstream bases; hence this proves that *en1FnCas9* has more tendency to create 3 base staggered lesions than other bases like 1-6 nucleotides (Figure 2.3).

### 2.3.2 Staggered DNA ends are processed differently than blunt ends in bacterial and mammalian cells

Although, it was previously assumed that *E. coli* merely depend on Homology-directed mechanisms to repair DSBs but integration of unrelated non-homologous sequences challenges this prior notion. It delves to delineate a mechanism termed A-EJ, culminating its differences from c-NHEJ (Wilson et al., 2003). A-EJ requires extensive end-resection fundamentally mediated by RecBCD complexes and leads to genome rearrangements, hence contributing to significant bacterial evolution. This mechanism

is not dependent on key NHEJ factors such as *Ligase-D* and *Ku* which are anyway absent in *E coli*, but A-EJ repair involves bidirectional resection, microhomology usage, and the involvement of specific proteins like *Ligase A*. Previous experimental setups include linear plasmid as substrates to evaluate the repair of DNA breaks in different mutagenic bacterial strains and explores the nature and efficiency of repair events in *E coli* strains (Chayot et al., 2010).

Unlike bacteria where NHEJ mechanisms are simple or absent, eukaryotes are proficient in NHEJ for repairing DSBs which are the potent drivers of genomic evolution. Mammalian cells have evolved complex machinery for NHEJ involving “core” components like *Ku70*, *Ku80*, DNA-dependent protein kinase, *Ligase4/XRCC4*, and other proteins that coordinate the repair process efficiently (Corneo et al., 2007; Guirouilh et al., 2007). DSBs created by CRISPR-Cas9 utilizes the cellular DNA repair pathways, including NHEJ, MMEJ, and HDR, to induce mutations at specific loci. Different mutational patterns produced by these pathways lead to semirandom indels by NHEJ, sequence specific indels by MMEJ and precise edits via HDR (McVey et al., 2008; Soulard et al., 2007; Yan et al., 2007). In the genome editing field, controlling the relative frequencies of these repair pathways for efficient and precise editing is the main intent.



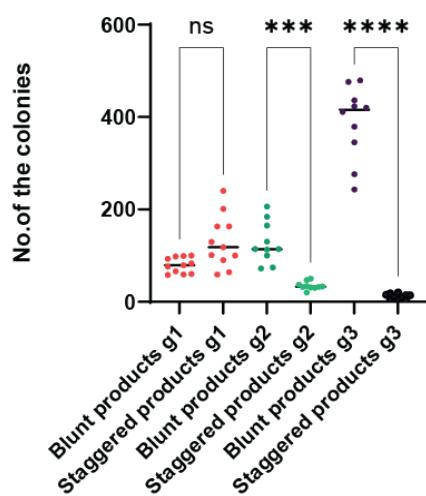
**Figure 2.4 Experimental paradigm of processing of *in-vitro* generated blunt and staggered DNA products by Cas9.** The linearized plasmid (open circle), is cleaved by SpCas9 and en1FnCas9 in *in-vitro* conditions targeting ampicillin region formed blunt and staggered linear DNA substrates and

transformed in *E. coli*, is repaired (recircularized) either accurately or inaccurately (Above panel). The linearized plasmid (open circle), is cleaved by SpCas9 and en1FnCas9 in in-vitro conditions targeting *Zsgreen* reporter region formed blunt and staggered linear DNA substrates and transfected in HEK293T cells, is repaired (recircularized) either accurately or inaccurately estimated by quantification of reporter expressing cells (Lower panel).

This sets the stage to investigate the editing outcomes of staggered (2-3 bases) and blunt DNA end products in *E. coli* which do not exhibit c-NHEJ activity and are commonly observed in other organisms. Additionally, this study aims to understand how different types of DSBs influence the existing A-EJ repair pathway. To examine this, we have generated blunt and staggered DSBs substrates cleaved by *SpCas9* and *en1 FnCas9* respectively targeting different ampicillin regions of pHIV *Zsgreen* plasmid. After DSBs formation, the intervening *Ampicillin* (Amp)- resistance gene is removed and these were delivered to *E. coli HST08* strain commercial competent cells to identify accurate repair events at blunt and staggered ends (Figure 2.4).

We observed that the efficiency of end-joining in *E. coli* is dependent on the substrate ends by counting the colonies in case of blunt and staggered DSBs. We found fewer colonies in staggered DSBs potentially because of initiation of resection, micro homologies and more possible number of in-frame deletions. We also calculated the tendency of MMEJ via scoring from a MMEJ prediction tool called RGEN and it is known that to avoid unwanted deletions the score should be more than **66**. From this method, we got score for **sgRNA 2 and sgRNA 3 as 59.5 and 61.8 (more than 66)** therefore, it has higher tendency of causing MMEJ, which could be the reason for more deletions during processing of staggered ends hence, we observed less colonies on plate for sgRNA 2 and sgRNA 3 (Figure 2.5).

<b>sgRNA1</b> CCCCGTCGTGTA <u>GATA</u> ACTACGGATACGGG <u>AGG</u> GCTTA CCCCGTCGTGTA <u>GATA</u> ACTAC <u>GATA</u> ACGGG <u>AGG</u> GCTTA CCC <u>CGTCG</u> TGTAGATAACTA <u>CGATA</u> ACGGG <u>AGG</u> GCTTA	<b>sgRNA 2</b> GGCATCGTGGTGT <u>TCACG</u> CTCGTCGTT <u>TGG</u> TATGG GGCATCGTGGT <u>GTC</u> ACG <u>GCTCGT</u> CGTT <u>TGG</u> TATGG GGCAT <u>CGTGGTGT</u> CAC <u>CGT</u> CGTT <u>TGG</u> TATGG	<b>sgRNA 3</b> GCT <u>CCT</u> TCGGT <u>CCTCCG</u> AT <u>CGT</u> GTCA GCT <u>CCT</u> TCGGT <u>CCTCCG</u> AT <u>CGT</u> GTCA GCT <u>CCT</u> <u>TCG</u> GT <u>CCCTCCG</u> AT <u>CGT</u> GTCA
---	---	--

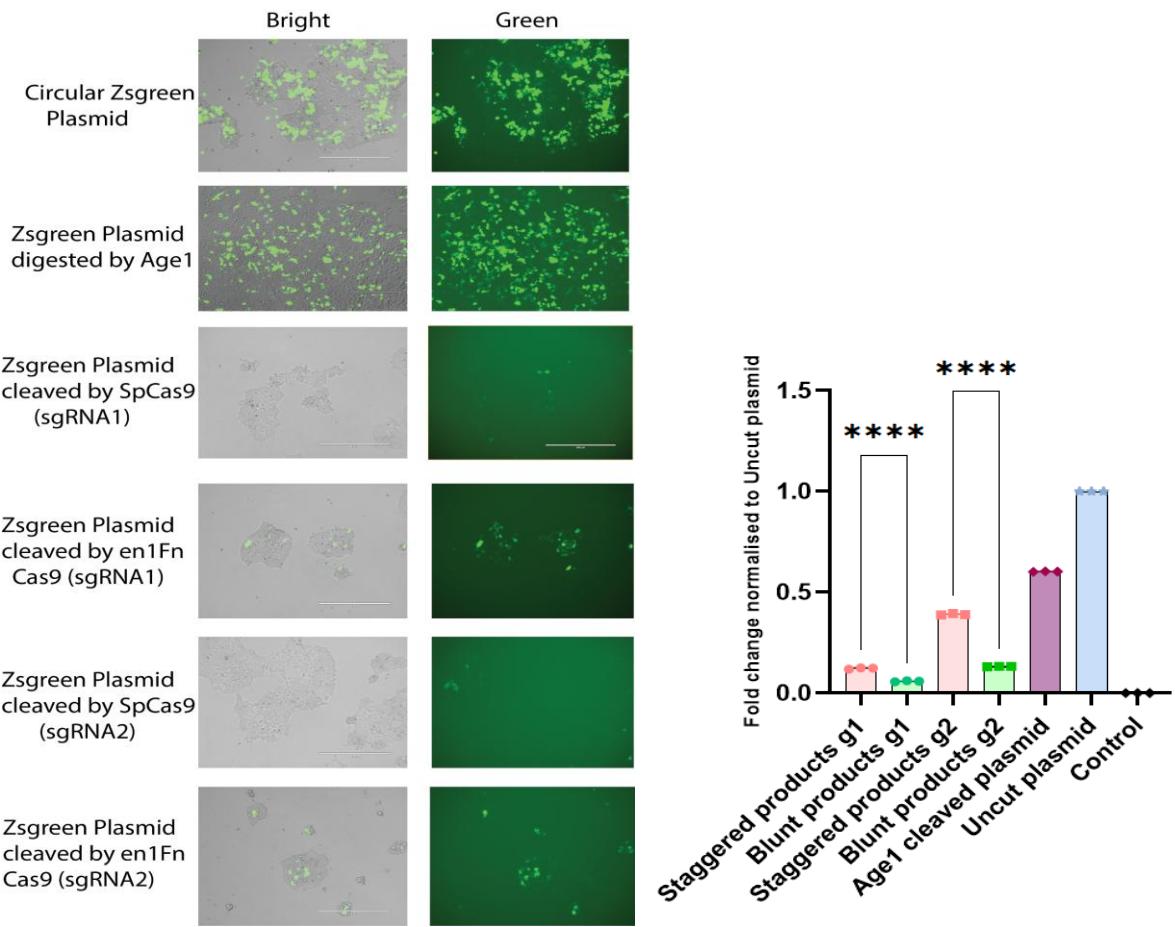


**Figure 2.5 Repair accuracy of blunt DNA ends is more than staggered DNA ends in bacterial cells.** (Left) Three guide RNA sequences sgRNA1, sgRNA2 and sgRNA3 in black bold targeting ampicillin and Micro Homologies (Green and Red) are highlighted and a yellow arrow indicates the PAM sequence in guide RNAs. (Right) Plot showing the no. of the colonies surviving in case of repairing blunt and 2-3 base staggered DNA substrates generated by three independent guide RNAs (g1, g2, g3). Error Bars represent SEM (10 independent experiments) and plotted on GraphPad prism. One-way anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001. (Data generated by Prabhleen Kuar).

In mammalian cells previous efforts like NHEJ inhibition, template design and its stabilization, cell synchronization for biasing the repair outcomes towards a desired pathway have already been established (Yang et al., 2020). Recently there is growing interest in understanding whether differences in double stranded structural breaks can influence the repair mechanisms and pathway preferences. DSB structure generated by natural and engineered Cas9 ranges from blunt ends to 1 base overhangs and staggered ends of longer lengths (>40 base pairs) have implications for increasing knock-in efficiency (Yang et al., 2020; Lemos et al., 2018; Shou et al., 2018).

Staggered bases of more than 1 base which we have observed in *en1 FnCas9* might have the potential to affect insertion and overall mutation efficiency. Distinct DNA cleavage patterns induced by Cas9 have prompted us to monitor the repair accuracy of blunt and staggered ended DNA substrates by transfecting it in mammalian cells targeting a reporter gene (Figure 2.4).

We have observed more accurate repair of *en1FnCas9* mediated staggered DNA ends than the *SpCas9* mediated blunt ends by measuring the Zsgreen expressing positive HEK293T cells. We have also compared this with Age1 restriction digested plasmids, which generates 4 base staggered DNA ends and act as the positive control for the assay (Figure 2.6).



**Figure 2.6 Repair accuracy of blunt DNA ends is less than staggered DNA ends in mammalian cells.** (Left Panel) Microscopic images showing the processing of blunt and staggered ends of DNA produced via SpCas9 and en1FnCas9 cleavage targeting *Zsgreen* reporter by two independent guide RNAs (g1 and g2) in HEK293T cells with scale bar 400um. (Right Panel) Bar plot showing the recircularization of *Zsgreen* plasmid depicting accurate repair outcomes represented as fold-change in FITC+ cells over the circular intact plasmid (normalized to Control cells). Error bars represent SEM (3 independent experiments) plotted on GraphPad prism. One-way anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001. Similarly, when we have calculated the score which is used to cleave the *Zsgreen* locus during *in-vitro* conditions (sg1 = 46.2 and sg2 = 49.8) it was less than 66, therefore tend to undergo MMEJ but still the accurate repair outcomes are observed during the repair of staggered ends potential leads to higher HDR. **Recent literature suggests that staggered DSBs can also be repaired and leads to accurate repair outcomes, but the factors that repairs these staggered DSBs are not well explored yet (Tov et al, 2024).**

## 2.4 Discussion

The DNA cleavage pattern typically results in 2–3 nucleotide overhangs at the 5' end of the non-target strand, a characteristic commonly observed with endonucleases such as Cpf1. These 5' overhangs are a defining feature of Cpf1's PAM-distal cleavage activity, where the cutting site is positioned farther from the PAM sequence. This generates sticky ends, which are single-stranded overhangs that exhibit reduced tolerance for mismatches in the DNA: RNA heteroduplex. This feature of Cpf1 contributes to its higher specificity and reduced off-target effects compared to other genome-editing tools like Cas9. In contrast, FnCas9/enFnCas9 produces PAM-proximal staggered cleavage at the 5' end of the non-target strand, a notable difference from Cpf1's PAM-distal cleavage activity. This divergence in cleavage patterns reflects distinct mechanisms of DNA processing, which can influence repair pathway choices and the precision of genome editing. The staggered cleavage observed with *FnCas9/enFnCas9* and *Cpf1* has significant implications for their use in genome-editing applications, as it can affect both the efficiency of editing and the likelihood of off-target effects. Understanding these cleavage patterns provides valuable insights for optimizing the design and application of CRISPR-based tools in various contexts.

For investigating the cleavage sites of engineered *FnCas9* (*enFnCas9*) variants across targets like *VEGFA3* and *HEK Site4* using the IVC assay, we have cloned Cas9 cleaved products that were sequenced by Sanger sequencing. As expected, we have observed that engineered *FnCas9* also nicks the target strand 3 bases upstream from the PAM, while the non-target strand exhibits variable nicking sites ranging from 4-7 bp away from the PAM similar to *FnCas9*. In contrast, *SpCas9*, as expected, nicks both strands 3 bp upstream from the PAM. Understanding these distinct cleavage patterns than the canonical Cas9 is crucial for supporting their application in targeted genome editing, minimizing off-target effects, and enhancing specificity. This detailed examination of cleavage patterns provides intriguing insights into the specificity and mechanisms of these CRISPR-associated proteins. The observed variability in nicking sites could have significant implications for their applications, especially in terms of targeting specificity and off-target effects. Investigating these differences could potentially lead to enhancements in genome editing precision and efficiency. Active engagement of researchers in controlling specific cellular repair pathways depending upon DSBs structures induced by Cas9 cleavage is crucial to ensure efficient and accurate modifications. Experimental setups from earlier studies have given the access to investigate the repair efficiency of blunt and staggered DSBs structures in bacterial and mammalian cells. We have noticed non-identical end processing of blunt and

staggered DNA substrates generated by *SpCas9* and *en1FnCas9* respectively. Nonetheless, regulatory mechanisms of repairing these DSBs are incredibly complex and yet to be comprehended entirely for revolutionizing the field of genetic medicine.

We pursued to discern the preference for nucleotide insertions induced by *SpCas9* and engineered variants of *FnCas9* *in-cellulo* conditions. We assessed insertion patterns in case of *SpCas9* and engineered variants of *FnCas9* targeting various endogenous locus. For this we have transfected the HEK293T cells with *SpCas9* and *en1FnCas9* containing plasmids with corresponding sgRNAs targeting different mammalian locus like *HBB*, *EMX1*, *FASN* and *LMNA C*. We sorted cells on the basis of transient GFP expression from Cas9-containing plasmids after 24 hours and performed the targeted amplicon sequencing on a Miseq platform.

**We found that the frequency of single nucleotide insertion in case of *SpCas9* and tri- nucleotides insertion in case of *en1FnCas9* mediated editing that depicts their staggered DNA ends under *in-cellulo* conditions.**

The single nucleotide insertion by SpCas9 was already observed in few of the recent reports, which also marks that it generates 1 bp DSBs. The finding of tri nucleotides insertion in case of en1FnCas9 mediated repair highlights the complexity of Cas9 cleavage mechanisms and their implications in genome editing.

## 2.5 Bibliography

1. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Biophysical Journal*, 106(2), p.695a.
2. Zhao, L., Bao, C., Shang, Y., He, X., Ma, C., Lei, X., Mi, D. and Sun, Y., 2020. The determinant of DNA repair pathway choices in ionising radiation-induced DNA double-strand breaks. *BioMed Research International*, 2020.
3. Shou, J., Li, J., Liu, Y. and Wu, Q., 2018. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Molecular cell*, 71(4), pp.498-509.
4. Kosicki, M., Tomberg, K. and Bradley, A., 2018. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nature biotechnology*, 36(8), pp.765-771.

5. Adikusuma, F., Piltz, S., Corbett, M.A., Turvey, M., McColl, S.R., Helbig, K.J., Beard, M.R., Hughes, J., Pomerantz, R.T. and Thomas, P.Q., 2018. Large deletions induced by Cas9 cleavage. *Nature*, 560(7717), pp. E8-E9.
6. Koralev, A., Lukyanchikova, V., Serova, I. and Battulin, N., 2020. On-target CRISPR/Cas9 activity can cause undesigned large deletion in mouse zygotes. *International Journal of Molecular Sciences*, 21(10), p.3604.
7. Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., Teichmann, M., Rousseau, E., Lamrissi-Garcia, I., Guyonnet-Duperat, V. and Bibeyran, A., 2019. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature communications*, 10(1), p.1136.
8. Zetsche, B., Abudayyeh, O.O., Gootenberg, J.S., Scott, D.A. and Zhang, F., 2020. A survey of genome editing activity for 16 Cas12a orthologs. *The Keio journal of medicine*, 69(3), pp.59-65.
9. Moreno-Mateos, M.A., Fernandez, J.P., Rouet, R., Vejnar, C.E., Lane, M.A., Mis, E., Khokha, M.K., Doudna, J.A. and Giraldez, A.J., 2017. CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature communications*, 8(1), p.2024.
10. Chayot, R., Montagne, B., Mazel, D. and Ricchetti, M., 2010. An end-joining repair mechanism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 107(5), pp.2141-2146.
11. Wilson, T.E., Topper, L.M. and Palmbos, P.L., 2003. Non-homologous end-joining: bacteria join the chromosome breakdance. *Trends in biochemical sciences*, 28(2), pp.62-66.
12. Corneo, B., Wendland, R.L., Deriano, L., Cui, X., Klein, I.A., Wong, S.Y., Arnal, S., Holub, A.J., Weller, G.R., Pancake, B.A. and Shah, S., 2007. Rag mutations reveal robust alternative end joining. *Nature*, 449(7161), pp.483-486.
13. Guirouilh-Barbat, J., Rass, E., Plo, I., Bertrand, P. and Lopez, B.S., 2007. Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proceedings of the National Academy of Sciences*, 104(52), pp.20902-20907.
14. McVey, M. and Lee, S.E., 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in Genetics*, 24(11), pp.529-538.
15. Soulard-Sprauel, P., Rivera-Munoz, P., Malivert, L., Le Guyader, G., Abramowski, V., Revy, P. and De Villartay, J.P., 2007. V (D) J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. *Oncogene*, 26(56), pp.7780-7791.
16. Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P. and Rajewsky, K., 2007. IgH class switching and translocations use a robust non-classical end-joining pathway. *Nature*, 449(7161), pp.478-482.

17. Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J. and Xia, N., 2020. Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. *International journal of molecular sciences*, 21(18), p.6461.
18. Lemos, B.R., Kaplan, A.C., Bae, J.E., Ferrazzoli, A.E., Kuo, J., Anand, R.P., Waterman, D.P. and Haber, J.E., 2018. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proceedings of the National Academy of Sciences*, 115(9), pp.E2040-E2047.
19. Shi, X., Shou, J., Mehryar, M.M., Li, J., Wang, L., Zhang, M., Huang, H., Sun, X. and Wu, Q., 2019. Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di-and tri-nucleotide CRISPR insertions without template donor. *Cell discovery*, 5(1), p.53.
20. Shou, J., Li, J., Liu, Y. and Wu, Q., 2018. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Molecular cell*, 71(4), pp.498-509.
21. Gisler, S., Gonçalves, J.P., Akhtar, W., de Jong, J., Pindyurin, A.V., Wessels, L.F. and van Lohuizen, M., 2019. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nature communications*, 10(1), p.1598.

## **Chapter 3: To comprehend the repair proteins recruitment triggered upon en1FnCas9 and SpCas9 DNA cleavage**

*This chapter focuses on investigating repair factors involved in staggered ended DNA repair via perturbations of essential repair related genes.*

### 3.1 Background

Indeed, the appreciation for repairing DNA DSBs is fundamentally ascribed to their importance in prolonging genomic DNA stability and cell survival. To address these DSBs, complex DNA repair mechanisms have been evolved by eukaryotic cells, essentially categorized into more prominent NHEJ and less prominent HDR (Shams et al., 2022). Besides, NHEJ and HDR constitute the major DSB repair pathways in eukaryotic cells, but the existence of repairing the DNA is far more complex than it is portrayed. Recent evidence also suggests the availability of hybrid DNA repair mechanisms that associate the NHEJ and HDR features. Understanding the complications of alternative DNA repair pathways and the regulatory networks governing their repair choice, is very important for underlying the involved mechanisms (Tang et al., 2019).

To overcome the technical challenges associated with the accurate profiling of DNA repair outcomes, an advancement has been developed known as Repair-seq (Chen et al., 2021). It involves the (CRISPRi)-based screens to provide the valuable insights of the specific repair factors participating in defined processing of the events due to simultaneous perturbations of thousands of genes. Its ability to assess the numerous genetic perturbations and repair outcomes concurrently, represents a notable improvement in the DSB repair study and genome editing technologies (Hussmann et al., 2021).

We have also comprehensively explored the specific factors responsible for staggered and blunt DSB processing via contemporaneous perturbations of thousands of genes. Our earlier studies were majorly in HEK293T cells, we have investigated the accurate sealing of blunt and staggered DSBs (generated by cleaving *Zsgreen* of pHIV *Zsgreen* mammalian expressing plasmid) followed by its transfection in dspCas9 KRAB expressing HEK293T that were simultaneously transduced with the CRISPRi gRNA library (Hussman et al., 2021). This library has 1,573 sgRNA targeting 476 essential genes that are repairing the DNA and associated with processes like DNA replication and recombination. The accurately sealed/repaired green positive and unsealed green negative cells were FACS sorted in both the conditions (staggered and blunt DSBs) and proceeded for single cell RNA sequencing, for identifying the specific factors involved in staggered ended DNA repair.

## 3.2 Methods and Materials

### 3.2.1 Plasmid Construction

PX408 (Addgene 68705) was used as a template for PCR amplification of the gene encoding the full length *Francisella novicida* (*FnCas9*) nuclease of residues 1–1629 bp as well as its modified variations. Following that, these genes were cloned using restriction enzyme-based cloning for the pET28-His-10-Smt3 vector (a gracious gift from Prof. Stewart Shuman and Dr. K.M. Sinha) and ligation-independent cloning (LIC) for the pET-His6-GFP-TEV-LIC vector (Addgene 29663), respectively. With minor changes to the manufacturer's procedure, Quick-change II site directed mutagenesis kit (Agilent) was used to create catalytically inactive *FnCas9* double mutants on pET His6-FnGFP-TEV-LIC plasmid backbone.

### 3.2.2 Protein purification

As stated earlier, the proteins employed in this investigation were purified. To put it briefly, *Escherichia coli* Rosetta2 (DE3) (Novagen) was used for creating plasmids encoding distinct Cas9 proteins. Until the OD 600 reached 0.6, the *Rosetta2 (DE3)* cells were grown in LB media supplemented with 50 mg/l kanamycin at 37°C. The addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) was used to stimulate protein expression. After an overnight cultivation at 18°C, the cells were separated by centrifugation. Re-suspended *E. Coli* cells were lysed by sonication and centrifugation in lysis buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol), supplemented with 1X protease inhibitor cocktail (PIC, Roche), and 100 µg/ml lysozyme. The lysate was affinity-purified by Ni-NTA beads from Roche, and the eluted protein was further purified using chromatography based on size-exclusion on a HiLoad Superdex 200 (16/60 column) from GE Healthcare in 150 mM KCl, 20 mM HEPES pH 7.5, 1 mM DTT and 10% glycerol. Thermo Fisher Scientific's Pierce BCA protein assay kit was used to quantify the concentration of purified proteins. Up until their next usage, the purified proteins were kept in storage at -80°C. For 6XHis-MBP-*dSpCas9*, *dFnCas9*, and its engineered variants the 6XHis-MBP was eliminated by soaking the affinity-bound protein in cleavage buffer (50 mM Tris-Cl, pH 8, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT) for a whole night in order to incubate the protein with PreScission Protease. On a HiLoad Superdex 200 16/60 column (GE Healthcare) with 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, and 1 mM DTT, the cleaved Cas9 protein was isolated from the fusion tag.

### **3.2.3 *In-vitro* transcription**

Using the Mega Script T7 kit (Thermo Fisher Scientific) and T7 promoters with templates as substrates, sgRNAs were transcribed in vitro. IVT reactions were purified using a NucAway spin column (Thermo Fisher Scientific) after being incubated for a full night at 37°C. Until they were needed again, IVT sgRNAs were kept at -20°C.

### **3.2.4 *In-vitro* Cleavage (IVC) assay**

For the DNA cleavage study, plasmid DNA containing the *Zsgreen* target sequence and the respective PAM (mentioned in respective legends) was used as the substrate for in vitro cleavage experiments. DNA template (50 ng or ~5 nM) was incubated at 37°C for 30 mins with the Cas9–sgRNA complex (50 nM) in 10 µL of reaction buffer, containing 20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, and 5% glycerol. The reaction was stopped by the addition of a quenching buffer, containing EDTA (20 mM final) and Proteinase K (40 ng).

### **3.2.5 Purification of the Cas9 cut linear products**

500 nM of Cas9-sgRNA complex was reconstituted at 25°C for 10 mins in reaction buffer with 500 ng of plasmid for *Zsgreen* locus. In vitro cleaved linearised plasmids were gel eluted using QI quick Gel extraction kit (Qiagen). These products were quantified by Qubit® dsDNA BR Assay as recommended by the manufacturer and an equal number of products were checked on agarose gel before transformation or transfection in cells.

### **3.2.6 Electroporation of AX227 pooled library in bacterial cells**

For achieving higher coverage (~1000 colonies/sgRNA), we have transformed 100 ng of pooled library into 60 ul of chilled MegaX cells, (2.5 kV, 4.9 ms in 0.2 cm cuvette). We have done the recovery in 1 mL total with SOC for 2 hours while shaking at 37°C. We have taken different dilutions of the recovered culture like 0, 200, 50 and 10 times and plated on the antibiotic plate. Rest of the recovery culture was added to 2000 mL LB+Carb. Grown O/N while shaking @ 37°C (16hrs). Transformation efficiency was calculated the following day and if it was higher than 1000 colonies per sgRNA construct in the library, then the cells were harvested and the library was purified by Qiagen maxi prep kit. Deep sequencing was done after amplifying and purifying the sgRNA containing region with the universal primers.

### **3.2.7 gRNA representation analysis in AX227 CRISPRi library**

To check for gRNA representation of the libraries, amplicons were generated from purified plasmid DNA. PCR primers were designed against conserved regions in the plasmid, forward primer was designed against the mU6 sequence and reverse primer was against the SpCas9 tracr sequence (forward primer: reverse primer:). Amplicon library was sequenced on Illumina NovaSeq 6000 with single-end 150 bp chemistry. Reads were demultiplexed on the instrument. Sequencing reads were quality-checked with FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.14 (<http://dx.doi.org/10.1093/bioinformatics/btw354>). Poor-quality bases with Phred score less than Q30 and TruSeq adapter sequences were removed using cutadapt v2.8 (<http://dx.doi.org/10.14806/ej.17.1.200>) for paired-end reads. Trimmed reads with at least 45 bp length were retained. Processed reads were aligned to respective libraries containing gRNA sequences. Uniquely mapped reads to respective gRNA sequences were counted and checked for distribution. To check for skewness in the library, lorenz plot was used with cumulative representation.

### **3.2.8 Cell culture and transfection**

Lenti HEK, HEK293T cells, dspKRAB HEK 293T cells were grown in DMEM media supplemented with high glucose (Invitrogen), 2 mM GlutaMax, 10% FBS (Invitrogen), 1X antibiotic and antimycotic (Invitrogen) at 37°C in 5% CO<sub>2</sub>. Transfections of mammalian cells were performed using Lipofectamine 3000 Reagent (Invitrogen) following the manufacturer's protocol.

### **3.2.9 Construction of the dspCas9 KRAB stable cell lines**

We then shifted our focus to explore the regulatory factors in repairing staggered and blunt ended DNA ends through perturbation experiments. For this purpose, we employed CRISPR-based methods to downregulate the essential repair gene expression in the HEK293T cells (CRISPR interference, CRISPRi) (Gilbert et al., 2014) ;(Alerasool et al., 2020) To address this we created dspCas9 KRAB stable cells and for this we have used addgene plasmid pHR-SFFV-KRAB-dCas9-P2A-mCherry for adequately expressing it via lentiviral viral based integration. These lenti *dspCas9 KRAB mcherry* constructs were used for the production of viral soup in Lenti HEK packaging cells with envelope and packaging plasmids and transduced it in HEK293T cells for the production of dspCas9 KRAB mcherry stable cell lines. Expression of dspCas9 KRAB in HEK293T cells were quantified by qRT PCR and Western Blot.

### **3.2.10 qRT-PCR**

Using the Qiagen RNeasy kit (74106) and the comprehensive instructions included in the kit, total RNA was extracted from stable cells expressing Cas9. After preparing cDNA synthesis using the Qiagen cDNA synthesis kit (205313) and following the recommended incubation conditions, the extracted RNA was treated with DNase I (Turbo DNase kit, Invitrogen, AM2238). Real-time qPCR was performed in triplicates on the test and control samples subsequent to cDNA synthesis. SYBR Green Master Mix: SYBR green based TB Green Premium Ex Taq II (Takara, RR82WR) was used in the PCR apparatus Light Cycler 480 (Roche) or Bio-Rad to quantify the transcripts. Every Ct value acquired for various transcripts was normalized using the GAPDH Ct value. Fold change analysis for comparative examination of the transcripts was done using the 2-DDCt method.

### **3.2.11 Western blot**

To create the cell protein lysate, the Cas9 stable cells were lysed in RIPA lysis buffer (ThermoFisher, Pierce). Each 6-well plate received 100 ul of the lysis solution and a dose of Protease Inhibitor Cocktail (PIC, Roche complete). For one hour, the cells were incubated at 4°C in a rocker. Each sample's protein lysate was obtained, and the Pierce BCA Protein Assay Kit (ThermoFisher) was used to assess the protein's concentration. An 8% SDS gel was loaded with 50 ug of protein for each sample, and PAGE was carried out using SDS running buffer (2.5 mM Tris base, 19 mM Glycine, and 0.1% SDS in autoclaved milliQ). The proteins were subsequently transferred utilising the Bio-Rad vertical gel transfer apparatus from the gel to the PVDF membrane (GE Healthcare Life-Science) using Transfer buffer (2.5 mM Tris base, 19 mM Glycine, 20% v/v Methanol in autoclaved milliQ) at 4°C for 3 hrs at 75 V. Following the completion of the transfer, the membrane was sliced to the appropriate protein size and stored for blocking with 5% BSA in 13 PBST (20 mM Tris base, 150 mM NaCl, and 0.2% Tween 20) at room temperature for two hours. The membrane was rocked. Following blocking, the blots were incubated for a further night at 4°C in a rocker with the primary antibody (Abcam 210752) at a 1:2500 dilution in the same blocking buffer. The loading control is a 1:1000 dilution of the vinculin antibody. Following the main antibody incubation, the blots were cleaned three times with 0.2% PBST for ten mins each. The identical blots were washed again, and then they were incubated for two hours at room temperature in a rocker with a secondary antibody that had an HRP conjugate. Upon the secondary antibody incubation, the blots had three 15-mins washes. The blots were developed using the Syngene Gel Doc instrument with EMD Millipore Immobilon Western Chemiluminescent HRP Substrate (ECL) to generate the signal. The

densitometry analysis and signal quantification were performed using ImageJ.

### **3.2.12 Immunofluorescence (IF)**

To prepare the slides, 22 x 22 mm coverslips (Corning, CLS285022) coated with fibronectin and gelatine were seeded with dspCas9 KRAB mcherry cells. Cells were seeded in a six-well plate at a density of approximately  $3 \times 10^5$  cells per well. After 24 hours, once the cells had adhered, they were collected by washing the wells twice with 1× PBS (Gibco™, 10010023). Since harvesting, every step of the process has been completed on well plates with cell-adhered coverslips. The cells were then fixed by incubating them at room temperature (RT) for 15 mins in a gentle shaker with cold 4% paraformaldehyde (PFA) + EDTA (Ph 7.4). The coverslips were twice rinsed with a washing buffer (0.1% Tween 20 in 1×PBS) after fixation. After incubating the cells for 10 mins at room temperature in a permeabilization solution containing 0.25% Triton X-100, the cells were twice washed with a washing buffer. Following permeabilization, the cells were treated for 60 mins at 37°C in a blocking solution containing 3% BSA in 0.1% PBS-Tween. After blocking, the cells were treated for an entire night with a 1:500 dilution of the Anti-HA primary antibody (Abcam ab9110) in a solution containing 1% BSA and 0.1% PBS-Tween. Three hours of incubation were spent at 37°C. By the conclusion of the initial incubation period, the cells were rinsed three times for five mins each using 1×PBST. Now, in order to precisely bind the primary antibody with the antibody, taking into account the species, the cells were treated with 488 Alexa fluor Anti Rabbit labelled fluorescent secondary antibodies (1:1000) dilution. (ThermoFisher Scientific) in the same blocking buffer and incubated for 60 mins at RT. The cells are rinsed three times with 1×PBST for five mins each after incubation. Following the washing, each coverslip received one drop of ProLongTM Diamond Antifade mountant with DAPI (InvitrogenTM, P369366). These coverslips were then mounted on one side of frosted glass slides (Corning, CLS294875X25) and imaging was carried out. The slides were visualized using a widefield fluorescence microscope, Leica TCS SP8 confocal microscopy equipment with 63X objective, and ThermoFisher Scientific's EVOS FL Auto imaging system. The additional table includes a list of the primary and secondary antibodies utilized, along with their dilutions.

### **3.2.13 CRISPRi dspCas9 KRAB HEK293T cells**

We selected the already published 1,573 sgRNA CRISPRi library (AX227), targeting 476 genes that are involved in DNA repair and associated processes (e.g., DNA replication, repair, recombination). It also contains 60 non-targeting control sgRNAs. Viral particles were produced in Lenti HEK cells

by co-transfected the AX227 library with packaging and envelope plasmids expressing gag/pol and envelope proteins respectively. Viral supernatant was syringed filtered of (0.45 µm) and titres were determined by Lenti-X™ qRT-PCR Titration Kit as per the manufacturer protocol. HEK 293T cells were transduced with AX227 library, infections were supplemented with 10 ug/ml of polybrene at MOI of less than 0.5. Cells were grown and selected with 2 ug/ml of puromycin (added 2-3 days post transduction). Cells were less than 50% BFP+ (sgRNA library marker) 2 days post transduction and for dead cells removal, cultures were washed periodically with PBS.

### **3.2.14 Uptake of exogenous templates by CRISPRi mammalian cells**

After quantification of the *in-vitro* Cas9 cleaved phiv *Zsgreen* plasmid products by Qubit, equal molecules of these blunt and staggered ended exogenous templates have been transfected in CRISPRi *dSpCas9 KRAB* HEK293T BFP positive cells via lipofectamine p3k reagent. Cells and ug of the products The GFP positive (accurately sealed) and GFP negative cells in both the staggered and blunt ended conditions were sorted after 24-48 hours of lipofection using BD FACS Melody Cell Sorter (BD Biosciences-US).

### **3.2.15 Repair-seq-sequencing library preparation**

Sample processing and library preparation- CRISPRi cells were processed using BD Rhapsody single cell analysis system as per Domenico et al and having certain modifications as described in Chattopadhyay et al (<https://doi.org/10.3389/fimmu.2022.1034159>). Briefly, 0.8 million cells per sample (5 samples) were taken and labelled using BD™ Single-Cell Multiplexing Kit-Human as per manufacturer's guide (Doc ID: 214419 Rev. 2.0). An average of  $4 \times 10^5$  cells were loaded in each cartridge on the BD Rhapsody express single cell analysis system for single cell capture. This was followed by the on-bead cDNA synthesis using BD Rhapsody™ WTA Amplification kit as per manufacturer's guideline (Doc ID: 210967 Rev. 1.0). The whole transcriptome amplification library was generated directly from the BD Rhapsody™ Enhanced Cell Capture Beads using a random priming approach, followed by an index PCR step. Meanwhile, the Sample Tag sequencing libraries, the extended Sample Tags were first denatured from the BD Rhapsody TM Enhanced Cell Capture Beads, which were later amplified through a series of PCR steps. The final libraries were quantified using Qubit high sensitivity DNA assay in Qubit 4 fluorometer, and the library size was determined using Agilent 4200 Tape station using high sensitivity DNA kit. Both the whole transcriptome mRNA and Sample Tag libraries were sequenced on NextSeq 2000 using NextSeq P3 sequencing kit at 30000 reads/cell for WTA, and 360 reads/cell/Sample Tag for sample tag library, with 85 x 215

read cycles configuration.

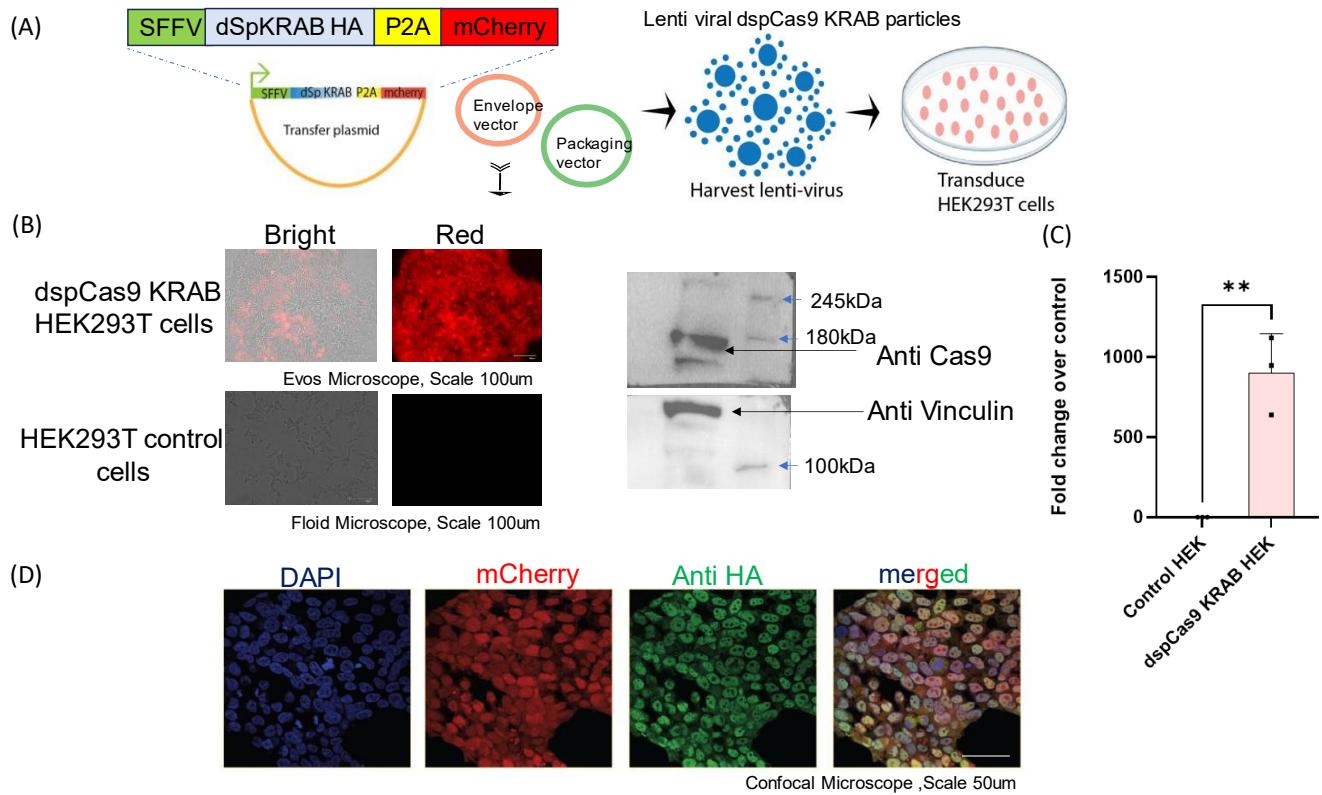
**scRNA-seq** data processing, clustering and differential gene expression- The raw sequencing data was converted to FASTQ format using the bcl2fastq tool. The data was analysed using BD Rhapsody WTA analysis pipeline as per manufacturer's guideline (Doc ID: 47383 Rev. 9.0). The count matrix with recursive substitution error correction was imported to the Seurat R package for downstream analysis and visualization. A total of 16,973 cells were detected post sequencing. Quality parameters were optimized and cells containing >8000 UMI and <200 UMI were discarded. A total of 12077 cells were obtained post QC, with an average of ~2400 cells per sample, which were taken for downstream analysis. Gene expression was normalized using the Log normalization within Seurat. Finally, the cells were clustered using unsupervised clustering at a resolution of 0.2 and visualized with tSNE algorithm. Cluster specific genes were identified using FindAllMarker function (Wilcoxon rank sum test, Log2 Fold Change cut-off 1.5). Data normalization was performed using the Log Normalize method within Seurat. Variable features were identified using the 'vst' method, selecting 2,000 features for downstream analysis. To ensure comprehensive analysis of target genes, all guide RNA-targeted genes were included in the variable feature set regardless of their variance status. Integration of multiple samples was performed using Seurat's standard integration workflow with conservative parameters as mentioned Anchor features: 2,000 plus all target genes; Dimensions: 1-20; k.filter: 50; k.score: 20; k.anchor: 5. Principal Component Analysis (PCA) was performed on the scaled data. Clustering was implemented using Seurat's FindNeighbors (dims = 1:20) and Find Clusters (resolution = 0.5) functions. UMAP dimensional reduction was applied for visualization using the first 20 principal components. Differential gene expression analysis was performed using the following parameters: Log2FC threshold- 0.1; Minimum percentage- 0.1; Statistical test- Wilcoxon rank sum test; Multiple testing correction- Benjamini-Hochberg; Minimum cells per feature- 3. Initial Expression Analysis was performed first to identify the downregulated genes in negative samples where repaired has not happened accurately in comparison with the control cells. Following this, status of expression for these corresponding genes was checked in the positive samples where repaired has happened accurately in both the staggered and blunt ends mediated repair. Since the gene expression was expected to differ in only 1 or 2 cells, log fold change was not considered, as the fold change is calculated as an average of all cells. Finally, the biological functions of the differentially expressed genes were identified using over representation analysis in EnrichR, with KEGG and Reactome database as the reference. Results were visualized using custom Python scripts implementing

matplotlib and seaborn libraries, creating enhanced volcano plots to display differential expression patterns and gene categories. Separate visualizations were generated for blunt and staggered ended repair conditions, incorporating various gene status categories and statistical significance thresholds.

### 3.3 Results

#### 3.3.1 Stably expressing dSpCas9 KRAB CRISPRi HEK293T cells

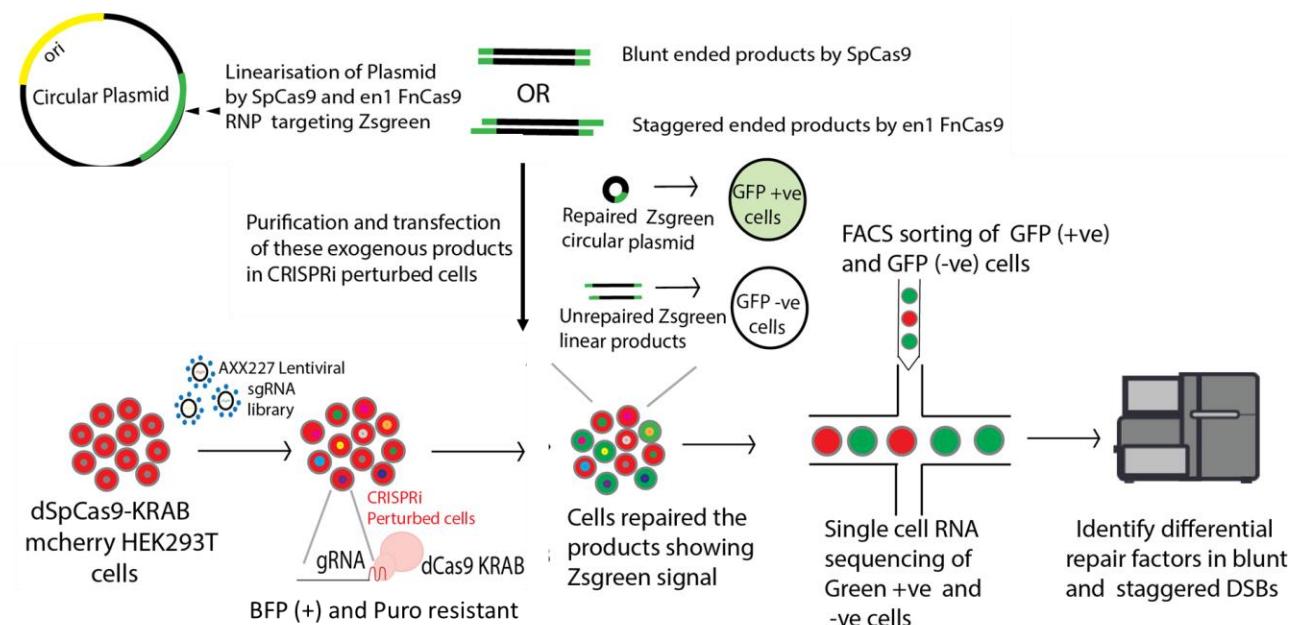
To understand the bias in pathway engagement for repairing staggered and blunt ended DSBs, we first created cells expressing dSpCas9 fused with a KRAB repressor in HEK293T cells via lenti-viral particles (Kampmann et al., 2018; Escriva et al., 2023). We have confirmed the RNA and protein expression of dSpCas9 KRAB via qRT PCR and Western blot, Immunofluorescence of single cells propagated clones (Figure 3.1).



**Figure 3.1 Generation of stable dSpCas9 KRAB expressing HEK293T cells.** (A) Schematic illustration of lentivirus-based overexpression of dSpCas9 KRAB in HEK293T cells. (B) Microscopic images (Left) with scale bar 100um and Western blot (Right) showing the protein expression of dSpCas9 KRAB in HEK293T cells with control HEK293T cells as 170kDa respectively. Vinculin taken as a loading control (117kDa). (C) Bar plots showing the RNA

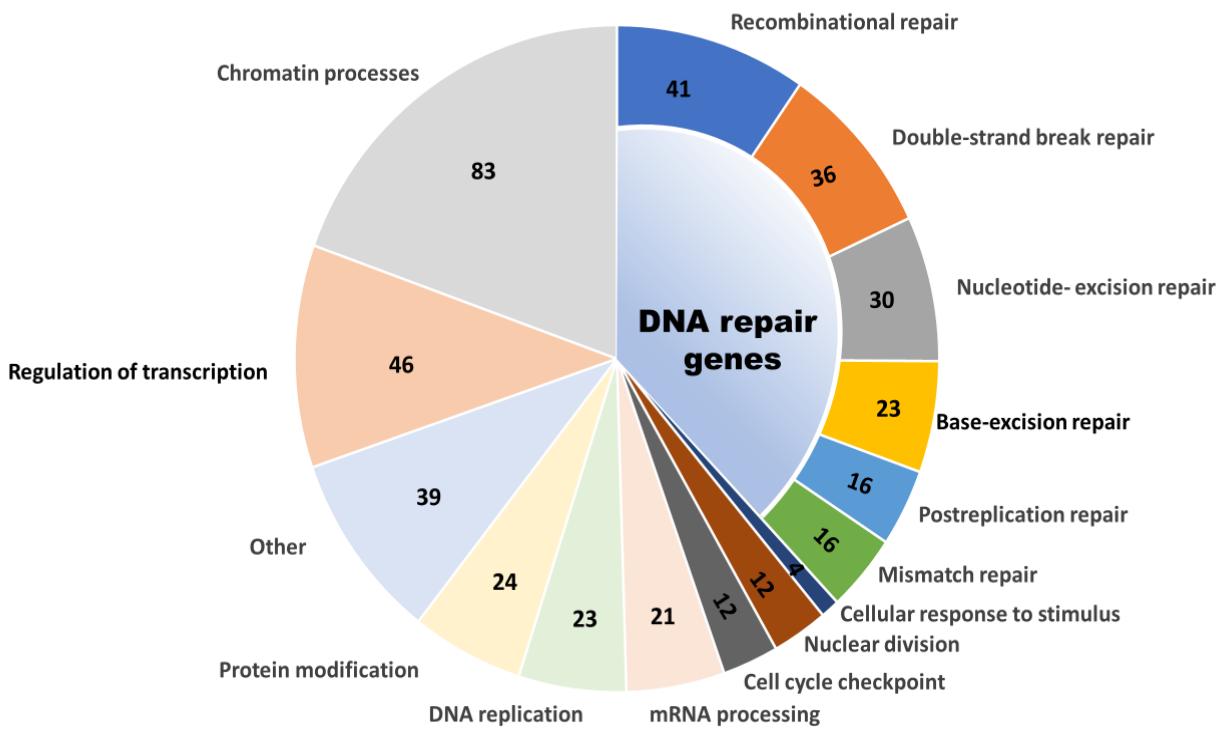
expression of dspCas9 KRAB in HEK293T cells estimated via qRT PCR. Error bars indicate the SEM of three independent replicates and plotted on GraphPad prism. Unpaired t test p values are shown \* $p<0.05$ , \*\* $p<0.01$ . (D) Representative Immunofluorescence images of dspCas9 KRAB tagged with HA (green) stained HEK293T cells that are also T2A mcherry positive. DNA is counterstained with DAPI and represented in blue. Scale bar 50 $\mu$ m.

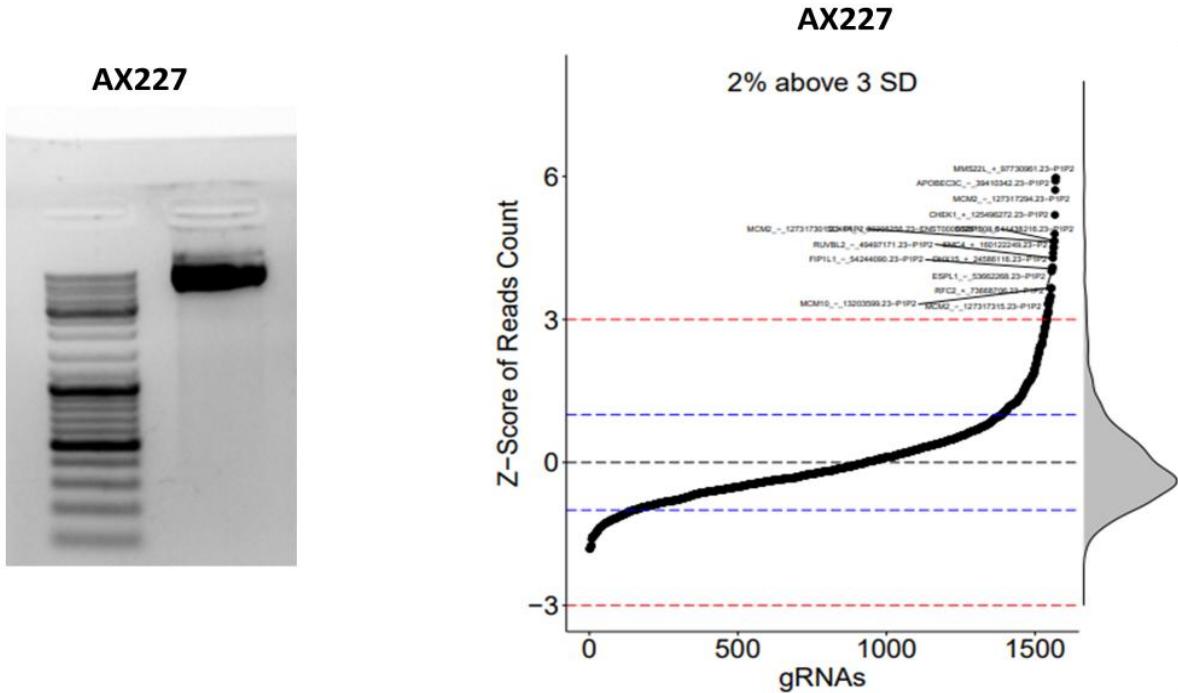
Following the expression profile of dSpCas9 KRAB in HEK293T cells, we have infected these cells with the CRISPRi library (AX227) adapted from the already published Hussman et al, Cell, 2021 with less than 1 multiplicity of viral infection (MOI). The well characterized dSpCas9 KRAB CRISPRi cells, were then transfected with the blunt and staggered DNA products, cells that accurately repair the staggered and the blunt products (green positive cells) were sorted from the un-repaired ones (green negative cells) and proceeded for single cell RNA sequencing (Figure 3.2).



**Figure 3.2 Experimental workflow for investigating repair dynamics induced upon SpCas9 and en1FnCas9 cleavage.** HEK293T cells stably express dCas9-KRAB effector protein (red colour), transduced with CRISPRi sgRNAs linked to DNA replication and repair process (Qi et al., 2013). Followed by adequate CRISPRi-mediated gene repression in HEK293T cells, *in-vitro* cleaved Zsgreen locus generated blunt and staggered end DNA products are delivered by lipofection. The FITC positive (Green Positive) and negative cells (Green Negative) containing the CRISPRi sgRNA and repair outcome is then FACS sorted and processed for single cell RNA sequencing to investigate perturbation-specific repair outcome distributions.

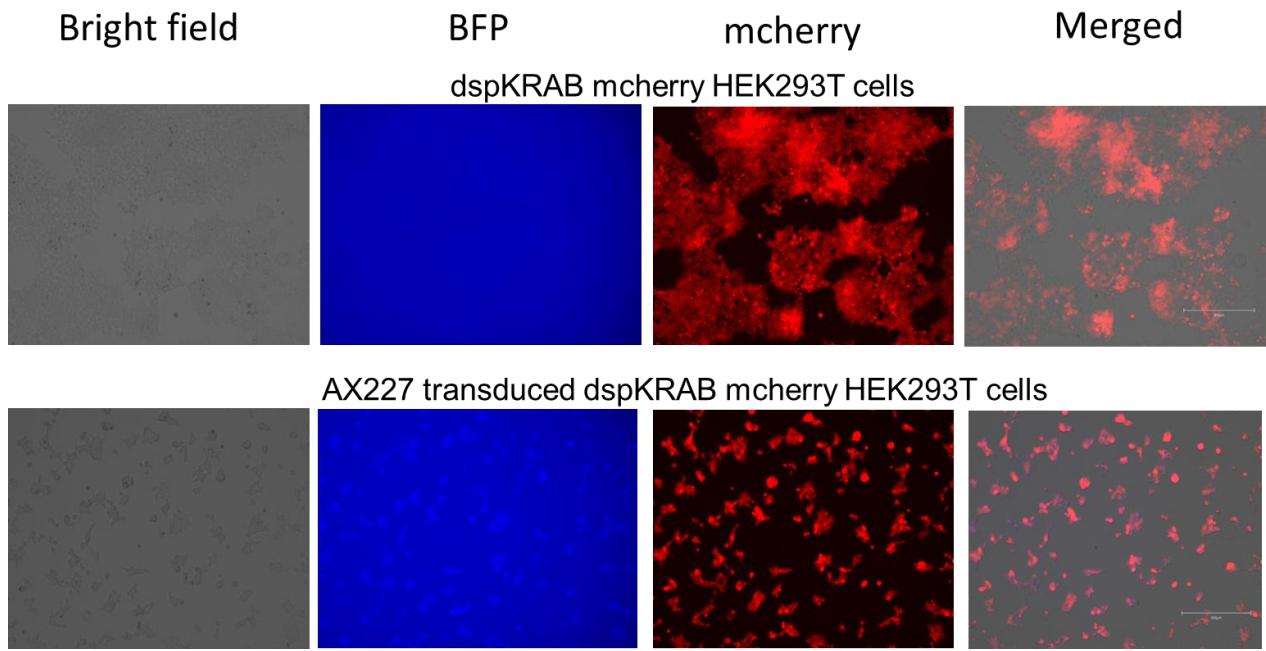
This library contained 1,573 sgRNA targeted 476 genes in DNA repair and associated processes like DNA repair, replication and recombination along with 60 non-targeting control sgRNAs. The arrangement of genes in AX227 library is such that 83 genes are responsible for chromatin process, 46 transcriptional regulation, 41 recombination repair, 36 DSB repair, 30 Nucleotide excision repair, 23 Base excision repair, 16 Post replication repair, 16 Mismatch repair, 4 cellular responses to stimulus, 12 Nuclear division, 12 cell cycle checkpoint, 21 mRNA processing, 23 DNA replication, 24 Protein modification, 39 others. Enrichment of the gRNA in pooled library with a selection (BFP) and survival (Puromycin) marker, was calculated by the deep sequencing of the guide RNA region by amplifying it with the universal primers. Z-score of the counts of the gRNAs were plotted and some have shown over representation such as *MMS22L*, *MCM2*, *MCM10*, *CHEK1*, *RUVBL2*, *FIP1L1*, *RFC2*, *ESPL1* and *APOBEC3C* (Figure 3.3).





**Figure 3.3 Characterisation of CRISPRi library (AX227) inhibiting DNA repair and associated processes.** (Upper Panel) Functional annotation of AX227 CRISPRi library comprising 476 genes targeted by the 1,573 sgRNA. (Lower Panel) Isolation of the pooled guide RNA containing plasmids in AX227 library and its gRNA representation analysis, Z scores of gRNA counts have been plotted.

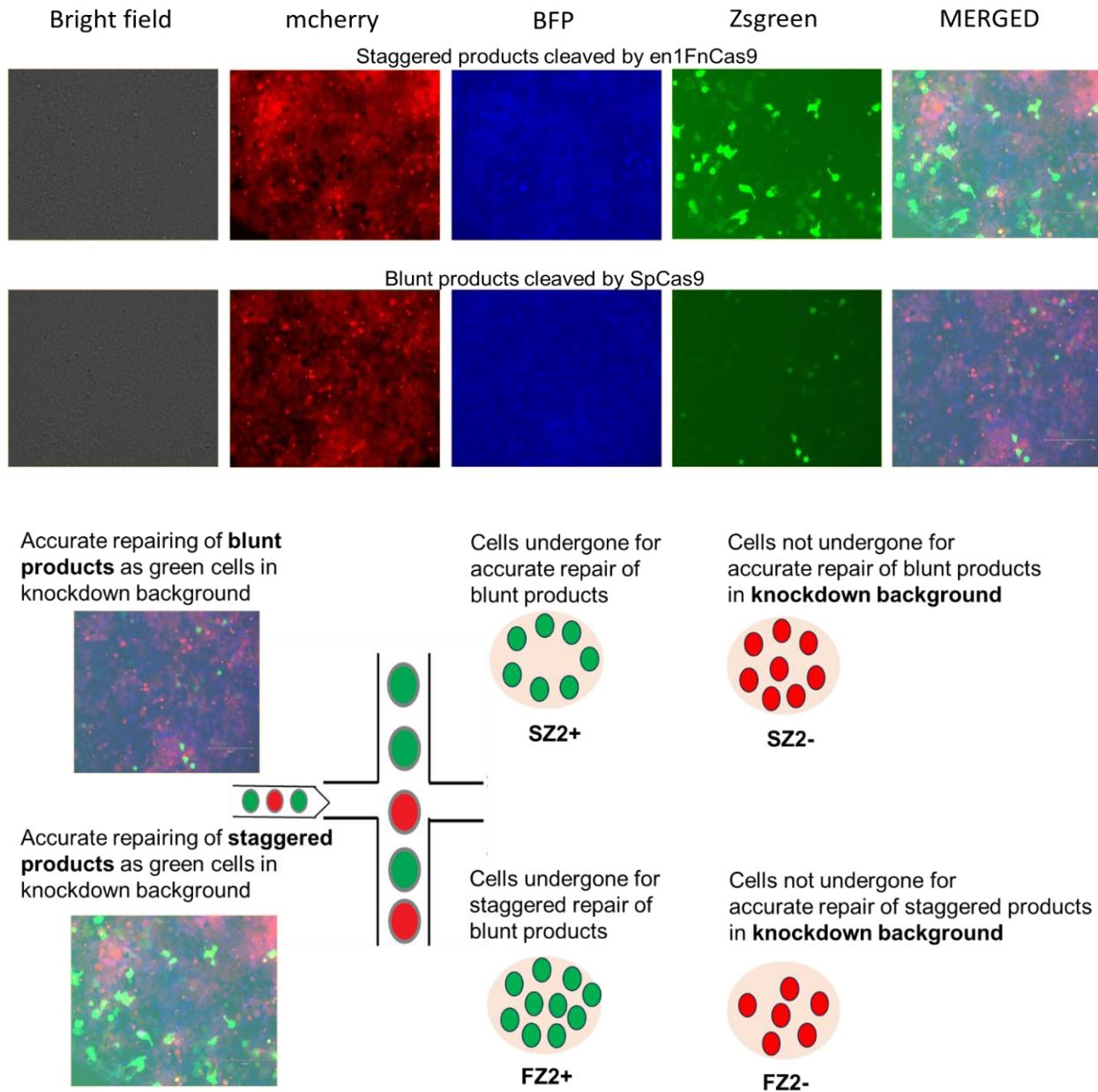
The AX227 library was transduced into HEK293T cells expressing dSpCas9-KRAB using lentiviral particles encapsulating pooled gRNA-containing plasmids (MOI < 1). Two days post-transduction, BFP-positive cells were observed (less than 50% of the cells were BFP-positive). These cells were then subjected to puromycin selection for 5–6 days. Following selection, cells were cultured until they reached 80–90% confluence, with dead cells removed via PBS washes. After 6 days of puromycin selection, all cells were BFP-positive due to successful integration of the AX227 library, and T2A-mCherry-positive, confirming stable expression of dSpCas9-KRAB in the HEK293T cells. (Figure 3.4).



**Figure 3.4 Microscopic images of untransduced and transduced BFP positive AX227 library in dspCas9 KRAB mcherry positive HEK293T cells.** Upper Panel showing the images of untransduced cells which are BFP negative and mcherry positive. Lower Panel showing the images of transduced cells which are BFP positive and mcherry positive after 6 days of puromycin selection. Scale bar 300 $\mu$ m.

### 3.3.2 Investigating differential blunt and staggered DNA repair factors via single cell RNA sequencing

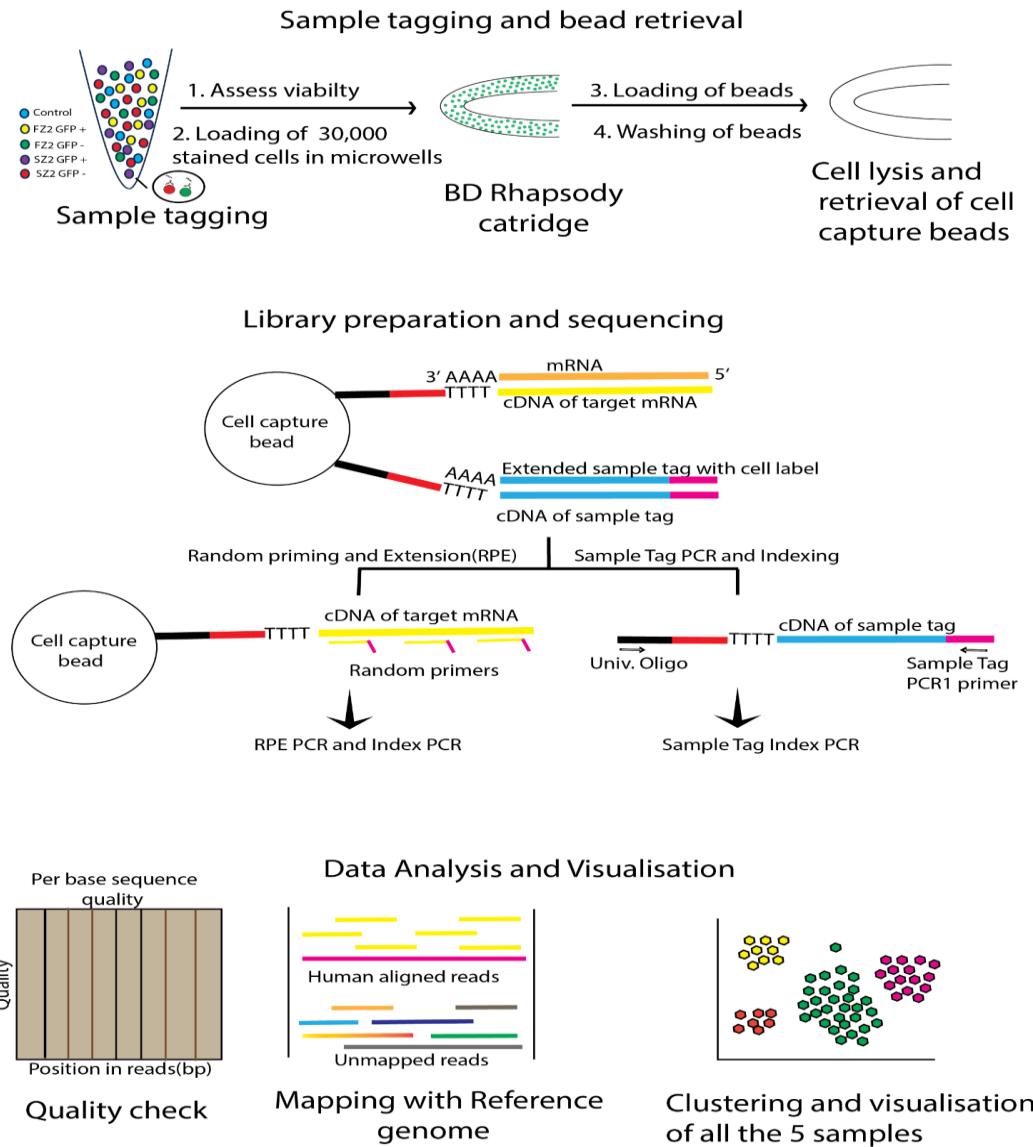
After allowing the repression of the genes involved in DNA repair, recombination and replication process in puromycin selection CRISPRi cells, we have transfected the *in-vitro* cleaved staggered and blunt ended products generated by *en1FnCas9* and *SpCas9* respectively targeting *Zsgreen*. To analyse the distribution of DNA repair outcomes caused by gene perturbations, we used FACS to isolate mCherry-positive + green-positive cells (accurately repaired) and mCherry-positive + green-negative cells (unrepaired). This was performed 24–48 hours post-transfection using the FITC channel, followed by single-cell RNA sequencing. (Figure 3.5). For single cell RNA sequencing, FACS sorted cells were proceeded for viability check and in all the samples like Control, green positive and negative cells in case of staggered ended repair outcomes annotated as **FZ2 + and FZ2 -**, green positive and negative cells in case of blunt ended repair outcomes denoted as **SZ2 + and SZ2-**.



**Figure 3.5 Single cell RNA sequencing of cells undergone for repairing staggered and blunt ended DNA products.** (Upper Pannel) Microscopic images of CRISPRi cells after 24 hours of transfection of staggered and blunt ended products generated by *en1FnCas9* and *SpCas9* respectively, targeting *Zsgreen* loci. First horizontal lane is showing the accurately and non-accurately repaired CRISPRi cells, transfected with the staggered products and second horizontal lane is showing the accurately and non-accurately repaired CRISPRi cells, transfected with the blunt products. Scale bar 150  $\mu$ m. (Lower Pannel) Experimental work flow of single cell sorting of cells undergone for accurate repair of blunt and staggered DNA ends, which are denoted as SZ2+ and FZ2+ green positive cells respectively. Cells which have not undergone for accurate repair of blunt

and staggered ends repair are shown as red cells and depicted as SZ2- and FZ2- respectively.

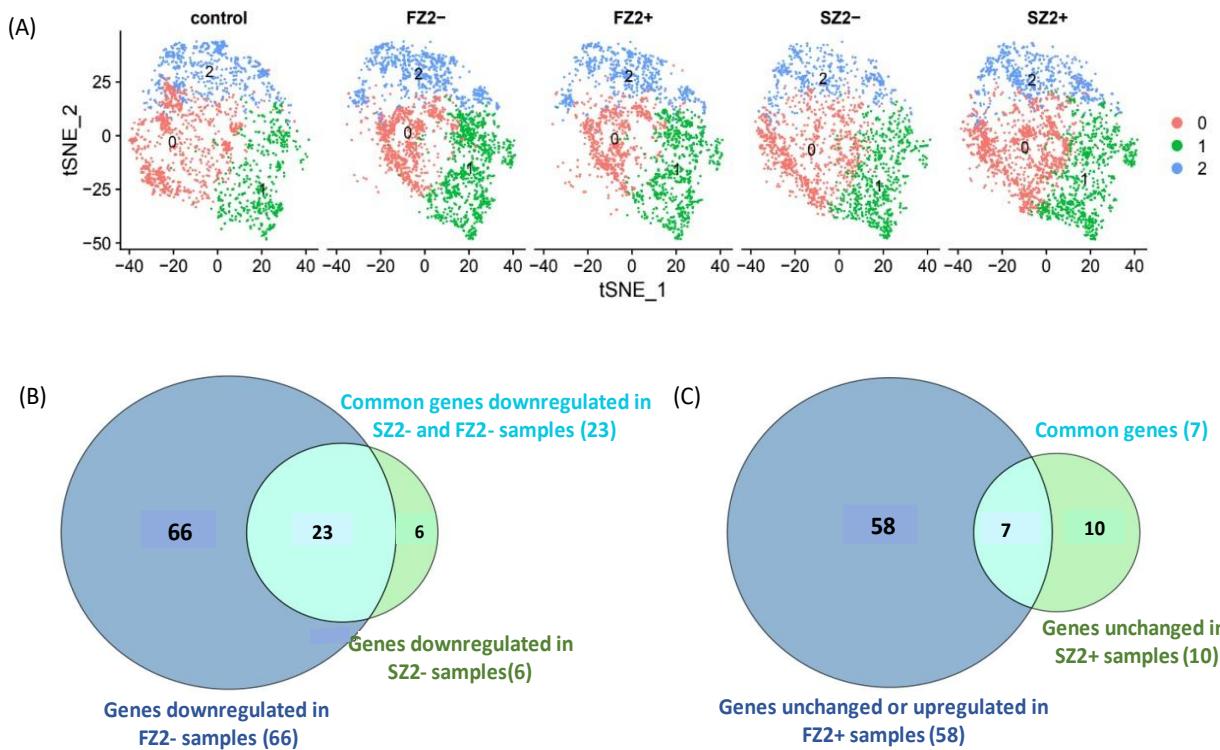
The viability in all the samples was more than 50 percent, and we got approximately 40, 961 cells after loading and the cell multiplet rate was 13.2%, the microwells are designed in such a way that it accommodates single cells coupled with the single bead. Subsequently, we proceeded with the bead loading and its washing according to the manufacturer's protocol. After bead washings, we performed the cell lysis step which involves the mRNA hybridisation to the beads. This is followed by the reverse transcription and amplification and library preparation as per the protocol (Figure 3.6).



**Figure 3.6 Pipeline for single cell RNA processing via BD Rhapsody platform.** Sample tagging and bead retrieval: After loading 30,000 cells into a microwell-based cartridge and labelling five samples—the control, FZ2 GFP+, FZ2 GFP-, SZ2 GFP+, and SZ2 GFP—individually with sample

tags, the beads were loaded and washed into the wells. Beads having mRNA linked to them are removed from the microwells following cell lysis. Library preparation and sequencing: The beads that contain mRNAs carrying Poly A are displayed with the mRNA capture sequences made up of oligo dT in front of the unique molecular index (UMI) in red and characteristic cell labels in black. Moreover, beads that catch the Poly A stretch sample tag from the same cell are shown in blue and magenta. On the beads, reverse transcription produces the cDNA for the sample tags as well as the mRNAs. Fragmented cDNAs are visible after barcoding, amplification, and indexing. data visualization and analysis.

Clustering analysis of the cells revealed three distinct clusters across all samples, including the control. Following the clustering based on gene expression, **Differential Expression Analysis** was conducted to identify genes involved in the repair of blunt and staggered DNA ends. Initially, the gene expression profiles of FZ2- and SZ2- conditions were compared to control samples. This analysis identified genes significantly downregulated in the negative conditions. The Venn diagram (Figure 3.7) illustrates the **unique (66 in FZ2- and 6 in SZ2-)** and overlapping (23) sets of downregulated genes in these conditions, highlighting their potential roles in DNA end repair mechanisms. Subsequently, a cross-comparison analysis was performed on positive samples (FZ2+ and SZ2+), which underwent accurate DNA end repair, in relative to their respective negative counterparts (FZ2- and SZ2-). This analysis compared the expression profiles of genes downregulated in FZ2- and SZ2- conditions with their counterparts in FZ2+ and SZ2+ samples. The comparisons between FZ2+ vs FZ2- and SZ2+ vs SZ2- identified genes that were either upregulated or remained unchanged in the positive samples. This revealed **distinct (58 in FZ2+ and 10 in SZ2+)** and shared (7) gene expression profiles between positive and negative conditions, providing valuable insights into genes potentially critical for repairing staggered and blunt DNA ends. (Figure 3.7).



**Figure 3.7 Quality control, Normalisation and Clustering of Single Cell RNA sequencing.** (A) T-distributed stochastic neighbor embedding (tSNE) plot shows clustering of 12077 cells from 5 samples (Control, FZ2-, FZ2+, SZ2-, SZ2+) and approximately 2400 cells per sample based on gene expression. Cell color specifies assignment of cells to 1 of 3 clusters (c0–2) inferred using shared nearest neighbor clustering. (B) Venn diagram depicting the gene expression analysis across FZ2- and SZ2- conditions compared to control samples, and showing the comparison of FZ2- vs Control and SZ2- vs Control, identifying downregulated genes in negative samples. (C) Venn diagram depicting the gene expression profiles of genes that were downregulated in FZ2- and SZ2- conditions (normalised with control), comparing them to their counterparts in FZ2+ and SZ2+ samples, and showing the cross comparison of FZ2+ vs FZ2- and SZ2+- vs SZ2-, identifying the upregulated or unchanged genes in FZ2+ and SZ2+ samples.

Among the distinct genes identified, 58 were specific to FZ2+ samples, while 10 were specific to SZ2+ samples. Of these, 5 genes were upregulated in FZ2+ samples but downregulated in FZ2- samples, indicating their potential roles in staggered DNA ends repair. The remaining 53 genes specific to FZ2+ and 10 genes specific to SZ2+ showed no significant change in expression

compared to control samples. However, these genes were downregulated in the negative samples, suggesting they may regulate DNA end-mediated repair processes. (Gene names are listed in Table H).

These genes were downregulated due to the targeted AX227 pooled CRISPRi guide RNAs screen, which specifically suppresses genes involved in DNA repair and associated processes such as DNA replication, recombination, nuclear division, and cell cycle checkpoints. The biological processes specifically linked to staggered and blunt DNA end repair are highlighted in the heat map presented in Figure 3.8. This analysis demonstrates that, in addition to **DNA repair**, **DNA recombination** emerges as a common pathway for both blunt and staggered end repair. Furthermore, processes like **DNA replication**, **DNA duplex unwinding**, **chromatin remodelling**, and **chromatin looping** are uniquely associated with staggered end repair. Together, these interconnected processes—duplex unwinding, chromatin remodelling, chromatin looping, and nucleosome assembly—form an integrated framework that facilitates accurate and efficient DNA repair, ensuring genomic stability and proper cellular function.

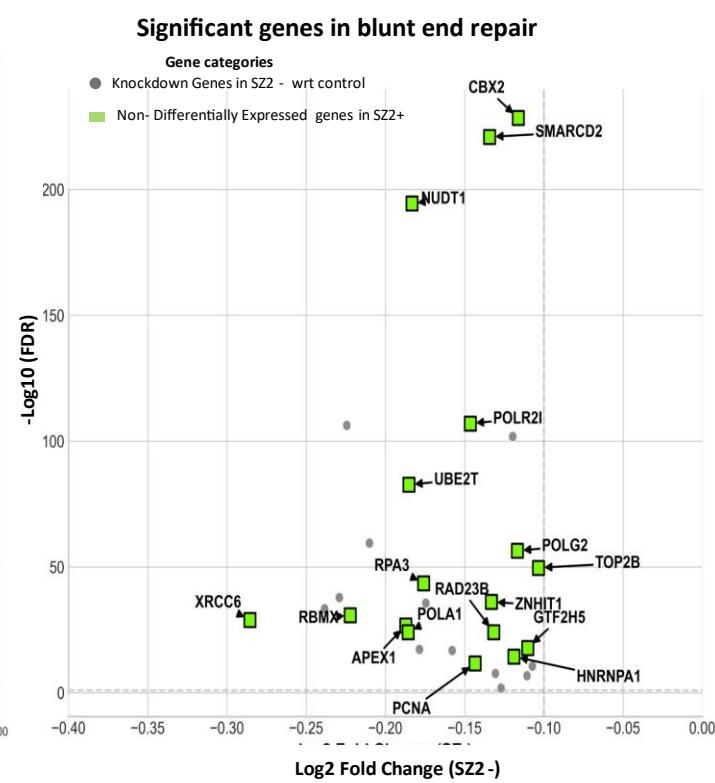
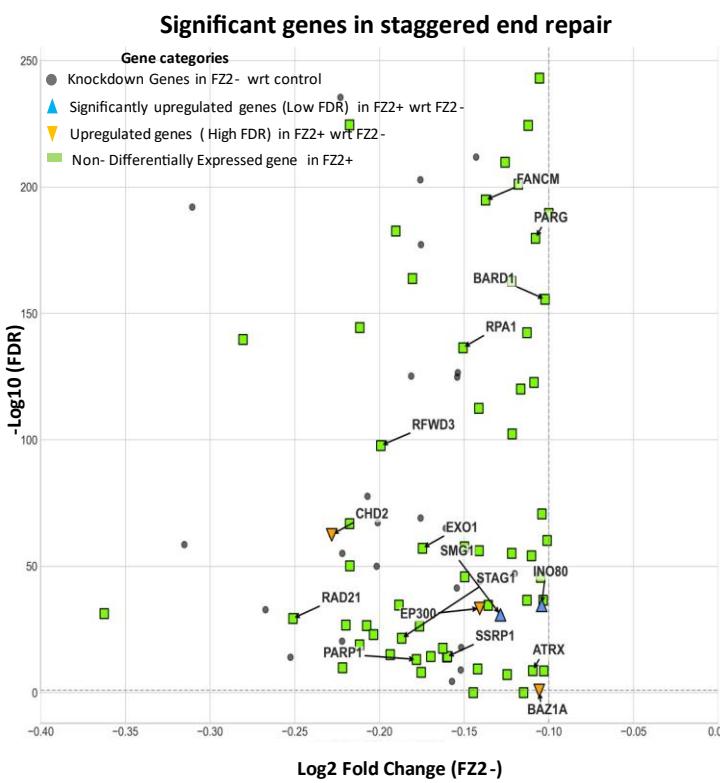
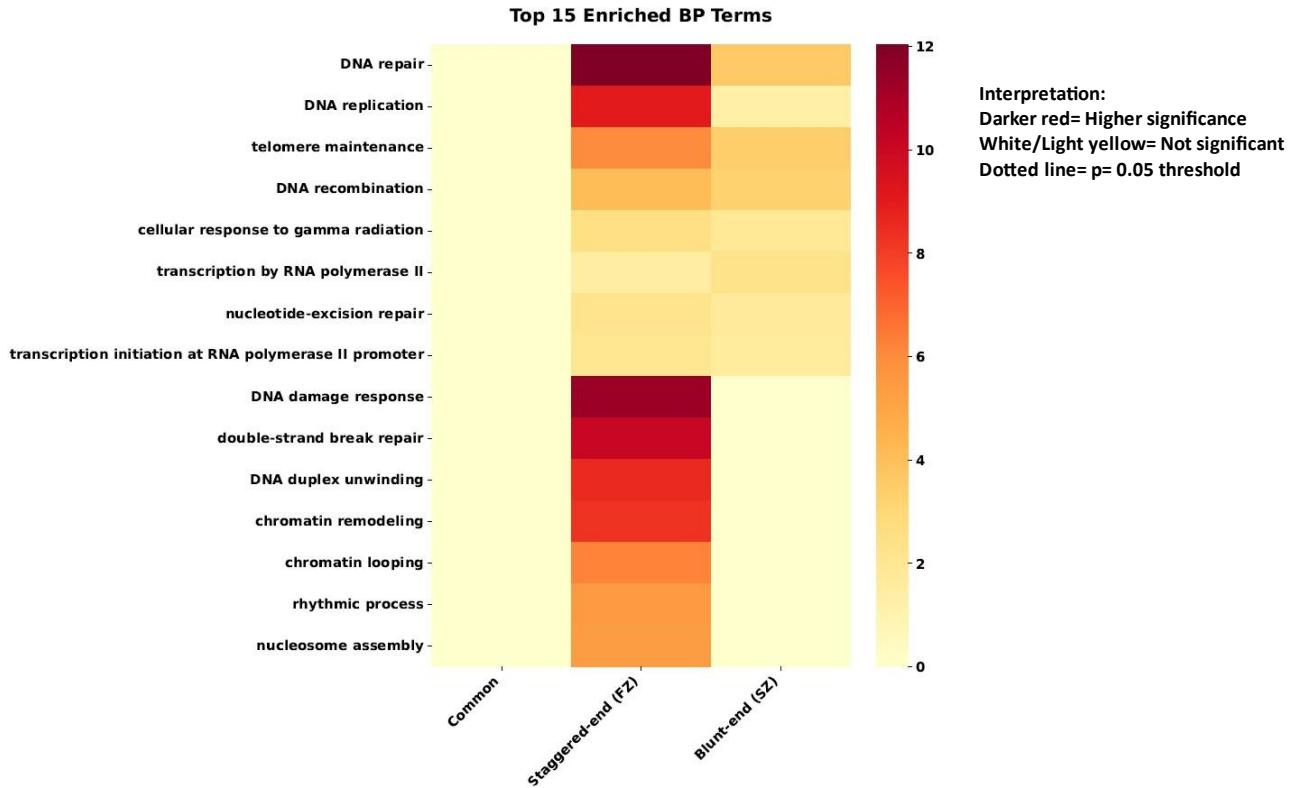
The genes identified as specific to **blunt end repair** includes- **XRCC6** is the core NHEJ component; binds DNA ends and recruits repair machinery, **APEX1** cleaves abasic sites in BER and acts as a redox coactivator, **CBX2** regulates DNA repair gene expression via chromatin remodelling, **GTF2H5** is essential for NER; unwinds DNA at sites of bulky lesions, **POLA1** synthesizes primers for DNA synthesis in HR and NHEJ, **POLR2I** links transcription to repair, particularly in TC-NER, **RBMX** promotes HR by facilitating RAD51 loading and genome stability, **RPA3** protects and stabilizes ssDNA intermediates during repair, **SMARCD2** facilitates chromatin remodelling at DNA damage sites, **UBE2T** which is the player in the FA pathway; monoubiquitinates FANCD2 and FANCI.

The genes identified as specific to **staggered end repair** includes **ALYREF** stabilizes R-loops and regulates DDR-related gene expression, **ARID1A** modulates chromatin for access to repair proteins; crucial for DSB repair, **ATRX** prevents replication fork collapse; involved in HR and telomere stability, **BARD1** partners with BRCA1 in HR repair and regulates DNA damage signalling, **BAZ1A** facilitates chromatin remodelling for repair by HR and NHEJ, **BCLAF1** regulates DNA damage-induced apoptosis and DDR signalling, **CBX6** maintains chromatin structure to support DNA repair processes, **CCNH** ensures repair of transcription-blocking lesions in transcription-coupled NER, **CDK7** coordinates transcription and NER; regulates cell cycle checkpoints for repair, **CHAF1A** restores chromatin structure post-repair; supports NER, **CHD2** facilitates chromatin access for DSB

repair and DDR gene activation, CHEK2 mediates cell cycle arrest and apoptosis in response to DSBs, DDX1 resolves R-loops to prevent replication stress and DNA breaks, DDX3X regulates DDR signalling and stabilizes replication forks, DICER1 processes small RNAs to regulate DDR gene expression; ensures genome stability, DNMT1 maintains DNA methylation; silences transposons to prevent mutagenesis, DROSHA processes miRNAs that regulate DDR-related gene expression, DTL prevents re-replication; coordinates DNA repair during the cell cycle, DUT provides dUTPase activity, preventing uracil incorporation into DNA during repair, EP300 functions as a histone acetyltransferase, promoting chromatin relaxation for repair, **EXO1** involved in DNA end resection during homologous recombination (HR) repair, **FANCM** facilitates DNA interstrand crosslink repair and replication fork stability, GTF2H1 component of the transcription factor TFIIH; critical for nucleotide excision repair (NER), GTF3C4 modulates transcription and repair processes, particularly RNA polymerase-dependent repair, HDAC2 deacetylates histones to regulate DNA damage response and chromatin structure, HIRA plays a role in histone replacement during repair, particularly in replication-associated repair, HNRNPUL1 facilitates DSB repair by regulating RNA splicing and DDR transcript processing, HP1BP3 maintains chromatin stability, indirectly supporting repair processes, **INO80** ATP-dependent chromatin remodeler critical for HR and replication fork stability, INTS3 part of the DNA-PK complex, supports telomere maintenance and replication stress response, LIG3 catalyzes DNA ligation during base excision repair (BER) and alternative NHEJ, MCM2 ensures replication fork progression and recovery during replication-associated repair, MTA2 modulates chromatin remodelling and repair, particularly during DDR, **PARG** removes PAR chains from PARP1 to regulate repair completion and prevent overactivation, **PARP1** detects single-strand breaks (SSBs) and recruits repair proteins for BER and other pathways, POLR3E supports transcription-coupled repair processes, PPP4R1 regulates DDR signalling by dephosphorylating key repair proteins, PRKDC encodes DNA-PKcs, a core component of NHEJ and DSB repair, **RAD21** cohesin subunit involved in sister chromatid cohesion and DSB repair by HR, RFC1 loads the PCNA clamp onto DNA during replication and repair, RFC2 part of the RFC complex, supporting PCNA loading and DNA repair synthesis, **RFWD3** promotes ubiquitination of RPA-coated DNA to facilitate HR repair, RNF2 modifies chromatin to recruit repair proteins at damage sites, **RPA1** binds to single-stranded DNA (ssDNA) and protects it during HR and replication stress, SMC4 involved in chromosome condensation and DNA repair through structural maintenance, SMG1 regulates nonsense-mediated decay and DDR signalling to ensure genome stability, **SSRP1** facilitates chromatin remodelling and transcription-associated repair, STAG1 cohesin subunit critical for sister chromatid cohesion and DSB repair by HR, SUMO2 mediates SUMOylation of repair proteins,

ensuring proper localization and function, SUZ12 part of the PRC2 complex, regulates chromatin state during DDR, TERF1 protects telomeres and prevents DNA damage signalling at chromosome ends, TOP1 resolves topological stress by relaxing supercoiled DNA during transcription and repair, TOP2A resolves DNA tangles and breaks during replication and repair, TP53 master regulator of DDR, promoting cell cycle arrest, apoptosis, or repair, TRRAP scaffolds chromatin remodelers to enable DNA damage signalling and repair, USP1 deubiquitinates repair proteins like FANCD2 to regulate repair pathways, XRCC5 part of the Ku heterodimer, essential for NHEJ-mediated DSB repair, YBX3b binds to DNA and RNA to stabilize damaged regions and support repair processes. The key genes for the staggered end repair includes **CHD2** which is a ATP-dependent chromatin remodeler; facilitates DNA access for repair, particularly in HR, **INO80** is a chromatin remodeler complex; promotes DNA end resection in HR and supports TCR in NER, **EP300** is a histone acetyltransferase; acetylates histones for chromatin relaxation, enabling DNA repair, **SMG1** is kinase involved in DDR and NHEJ; regulates checkpoint signalling and mRNA decay related to DNA repair, **BAZ1A** is also ATP-dependent chromatin remodelling; aids in HR repair by modifying chromatin structure at DSBs.

Given their functional significance, these genes are strong candidates for further exploration of their roles in staggered (notably, the highlighted genes have already been reported to facilitate homologous recombination pathways) and blunt DNA end repair. However, additional experimental validation is necessary to confirm their specific functions and contributions to these repair mechanisms. (Figure 3.8).

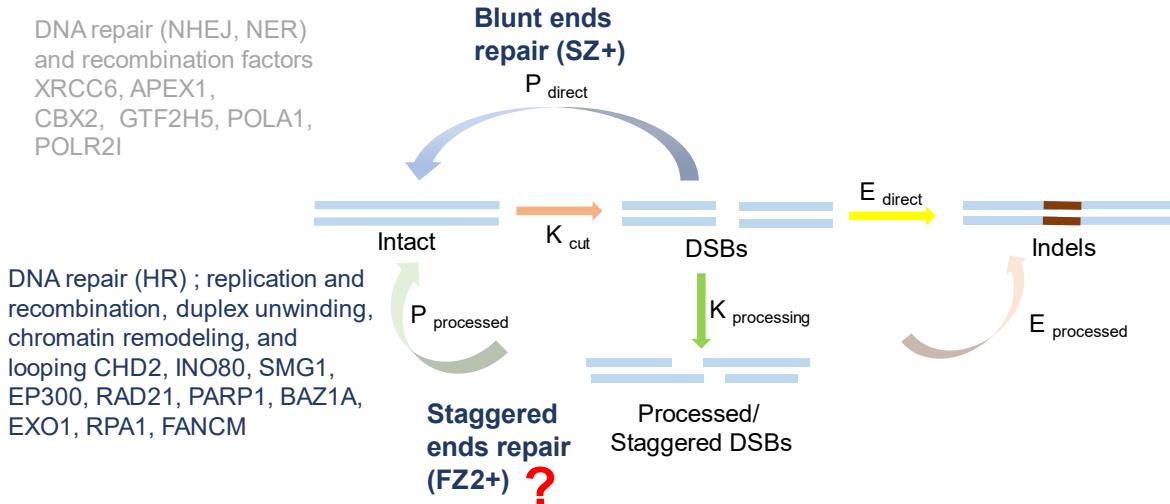


**Figure 3.8 Differential Expression Analysis of blunt and staggered ends DNA repair genes.**

**(Upper Pannel of Heat Map)** Gene Ontology enrichment analysis of biological processes in common pattern genes and structure-specific genes. Heatmap shows the top 15 enriched biological processes (BP) based on significance (-log10 adjusted p-value). The color intensity represents statistical significance, with darker red indicating higher significance (lower p-value) and light yellow indicating lower significance. The three columns represent common genes (Common), staggered-end specific genes (FZ), and blunt-end specific genes (SZ). The dotted line on the color scale indicates the significance threshold ( $p = 0.05$ ). Analysis was performed using gene sets derived from RNA-seq data, with Benjamini-Hochberg correction for multiple testing.

**(Lower Pannel of Volcano Plots)** (Left volcano plot) Differential expression analysis of genes in blunt-end negative (SZ-) versus positive (SZ+) conditions. Volcano plot showing the relationship between statistical significance (-log10(FDR)) and magnitude of change (log2 fold change) for guide genes. Green squares represent genes that show no significant change in blunt-end positive (SZ+) samples, while grey circles indicate knockdown genes. Gene symbols are shown for key differentially expressed genes. The vertical dotted line represents the significance threshold. The x-axis shows log2 fold change in SZ- condition, and the y-axis shows the statistical significance as -log10(FDR). (Right volcano plot) Differential expression analysis of genes in staggered-end negative (FZ-) versus positive (FZ+) conditions. Volcano plot showing the relationship between statistical significance (-log10(FDR)) and magnitude of change (log2 fold change) for guide genes. Green squares represent non-differentially expressed genes in staggered-end positive (FZ+), blue triangles indicate significantly upregulated genes (low FDR), yellow triangles show upregulated genes (high FDR), and grey circles indicate knockdown genes. Gene symbols are labelled for key differentially expressed genes. The vertical dotted line represents the significance threshold. The x-axis shows log2 fold change in FZ- condition, and the y-axis shows the statistical significance as -log10 (FDR). Heat maps and Volcano plots have been generated by Vishal Bharti.

This highlights that the accurate repair of staggered double-strand breaks (DSBs) is primarily mediated by homologous recombination (HR) proteins (highlighted in the section above) and further supported by processes such as DNA replication, DNA duplex unwinding, chromatin remodelling, and chromatin looping (Figure 3.8). This observation led us to investigate whether 2-3 base staggered DNA ends are sufficient to enhance the HDR/NHEJ ratio in cells when using HDR templates of varying lengths and types.



71

**Figure 3.9 Schematic summarizing the processing of staggered DSB induction and its repair.**

Kcut, ends processing, Kprocessing, repair from directly induced DSBs, namely precise repair,  $P_{\text{direct}}$  mainly by NHEJ, MMR, NER promoting factors contributing SZ<sup>+</sup> and processed DSBs, including precise repair  $P_{\text{processed}}$  via HDR promoting factors contributing FZ2<sup>+</sup>.

### 3.4 Discussion

The advancement of genome engineering tools like CRISPR-Cas9 and our understanding of genome stability depend on our capacity to analyse DNA DSBs and their repair processes. Due to imprecise repair, DNA DSBs pose a serious danger to genome integrity and promote genome evolution. Genome engineering has been revolutionized by the ability to target DSBs to certain sequences using specifically created nucleases, particularly CRISPR-Cas9. A complicated web of paths leads to indels, or conversion tracts from templated repair when DSBs are repaired.

These mechanisms may be roughly classified into two categories: homologous recombination (HR) and non-homologous end-joining (NHEJ). In mammalian cells, the key mechanism is canonical NHEJ (cNHEJ), however other end-joining (alt-EJ) pathways, such Polymerase Theta Mediated End-Joining (TMEJ) and Microhomology-mediated end-joining (MMEJ), can also function. At the junctions, random integrations (RI) mediated by NHEJ are distinguished by  $\leq 2$  base pair (bp)

homology. TMEJ-mediated RI usually involves significant terminal deletions and junctional insertions along with 2 to 6 bp of microhomology between the foreign DNA and the genome. Research indicates that NHEJ may be quite exact in mammalian systems, with precise repair happening in mouse cells as much as 75 % of the time. The main areas of study for CRISPR/Cas9-induced double-strand breaks in endogenous chromatin are mutagenesis repair because of random integration (RI) and variables influencing the choice of repair pathway.

Although less common than RI, targeted integration (TI) arises when foreign DNA has homologous areas with the genome, rendering gene-targeting technologies ineffective for optimal gene therapy. In addition to enhancing TI, CRISPR-Cas9-mediated DSB induction at the target locus also induces RI at both on- and off-target locations. It can result in chromosomal rearrangements or inter-homologous recombination to repair DSBs. Cells with Rad51 deletion retain the ability to target genes, albeit at a reduced frequency, indicating the presence of a non-canonical, Rad51-independent pathway. When there are defects in other DSB repair mechanisms, this noncanonical gene targeting mechanism becomes more prominent and is reduced by ***Rad52 suppression***. *Rad51*-independent gene targeting appears to function at many phases of the cell cycle, as it is not limited to the S-G2 phase.

Although kinetic investigations shed light on the mechanics of DSB induction and error-prone repair, it is still difficult to measure accurate repair and see scar-less re-ligation firsthand. Although it is already known that NHEJ is mostly used for the scarless ligation in case of blunt DSBs, research has not yet been done on the route processed DSBs, which are already referred to as staggered DSBs created by Cas9 in the literature. Our results demonstrate that the repair of 3-4 base 5' staggered double-strand breaks (DSBs) generated by en1FnCas9 is primarily facilitated by factors that favor homologous recombination (HDR), notably **CHD2**, **INO80**, **EP300**, and **BAZ1A**. These key proteins play a crucial role in ensuring that the DNA repair process proceeds efficiently and accurately.

However, DNA repair in case of staggered ends is not solely reliant on these proteins; it is observed to be a multifaceted process that also involves several other essential factors. **DNA replication** helps provide the necessary templates for repair and ensures that the repaired DNA is correctly integrated into the genome. **Duplex unwinding** is critical for exposing the damaged DNA regions, allowing repair proteins to access the break sites. **Chromatin remodelling** and **chromatin looping** further enhance the accessibility of the damaged DNA by modifying the chromatin structure, facilitating the recruitment of repair complexes. Additionally, **nucleosome assembly** ensures that the chromatin

structure is properly restored after repair, maintaining genomic stability.

Collectively, these processes—HDR repair proteins, DNA replication, duplex unwinding, chromatin remodelling, looping, and nucleosome assembly—work in concert to ensure efficient and precise DNA repair. Their coordinated action enables the efficient recruitment, stabilization, and function of repair complexes, ultimately restoring the chromatin structure and maintaining the integrity of the genome after damage. This intricate network of repair processes highlights the complexity and precision required for accurate DNA repair mechanisms.

Important new understandings of the intricate interactions between various DNA repair processes are provided by research on DNA integration mechanisms in mammalian cells. The knowledge acquired may contribute to the improvement of gene-editing technologies, increasing their effectiveness and dependability for medicinal uses. By dissecting these pathways and mechanisms, the study enhances our comprehension of DNA repair processes and their implications for gene therapy, offering potential strategies to improve gene-targeting techniques.

### 3.5 Bibliography

1. Tang, X.D., Gao, F., Liu, M.J., Fan, Q.L., Chen, D.K. and Ma, W.T., 2019. Methods for enhancing clustered regularly interspaced short palindromic repeats/Cas9-mediated homology-directed repair efficiency. *Frontiers in genetics*, 10, p.551.
2. Hussmann, J.A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., Simpson, D., Yang, D., Bothmer, A., Cotta-Ramusino, C. and Weissman, J.S., 2021. Mapping the genetic landscape of DNA double-strand break repair. *Cell*, 184(22), pp.5653-5669.
3. Chen, P.J., Hussmann, J.A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.F., Chen, C., Nelson, J.W., Newby, G.A., Sahin, M. and Osborn, M.J., 2021. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell*, 184(22), pp.5635-5652.
4. Shams, F., Bayat, H., Mohammadian, O., Mahboudi, S., Vahidnezhad, H., Soosanabadi, M. and Rahimpour, A., 2022. Advance trends in targeting homology-directed repair for accurate gene editing: An inclusive review of small molecules and modified CRISPR-Cas9 systems. *BioImpacts: BI*, 12(4), p.371.
5. Kampmann, M., 2018. CRISPRi and CRISPRa screens in mammalian cells for precision biology and

- medicine. *ACS chemical biology*, 13(2), pp.406-416.
- 6. Escrivá-Fernández, J., Cueto-Ureña, C., Solana-Orts, A., Lledó, E., Ballester-Lurbe, B. and Poch, E., 2023. A CRISPR interference strategy for gene expression silencing in multiple myeloma cell lines. *Journal of Biological Engineering*, 17(1), p.34.
  - 7. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5), pp.1173-1183.
  - 8. Ben-Tov, D., Mafessoni, F., Cucuy, A., Honig, A., Melamed-Bessudo, C. and Levy, A.A., 2024. Uncovering the dynamics of precise repair at CRISPR/Cas9-induced double-strand breaks. *Nature Communications*, 15(1), p.5096.
  - 9. Saito, S. and Adachi, N., 2024. Characterization and regulation of cell cycle-independent noncanonical gene targeting. *Nature Communications*, 15(1), pp.1-13.
  - 10. Pavani R, Tripathi V, Vrtis K, Zong D, Chari R, Callen E, Pankajam K, Zhen G, Yang J, Wu W, Cejka P, Walter J, Nussenzweig., 2024. Structure and repair of replication-coupled DNA breaks. *Science*.

## **Chapter 4: To examine the repair outcomes mediated by en1 FnCas9 and SpCas9 at different loci.**

*Summary of how FnCas9 and its engineered variant 2-3 base staggered cutting mechanism induce higher HDR/Indels ratio than blunt ended DNA in different cell lines.*

## 4.1 Background

Utterly, the favourable outcomes of the CRISPR-Cas system are indeed intertwined with intricate DNA repair mechanisms. The four major pathways for repairing DNA DSBs are crucial to comprehend when utilizing CRISPR Cas as genome editing tool and greatly influences the accuracy and efficiency of its editing (Peng et al., 2014; Chang et al., 2017; Salsman et al., 2017; Dumitracne et al., 2011; Blackford et al., 2017; Sulli et al., 2012; Schwertman et al., 2016). One of the major DNA repair pathways **NHEJ** is error-prone in nature as it directly rejoins the fragmented DNA ends and results in small indels at the Double stranded break DNA ends (DSBs) (Fell et al., 2015; Kragelund et al., 2016; Davis et al., 2014; Conlin et al., 2017; Symington et al., 2016). The highly efficient NHEJ repair factors employed by the cells can bring in undesired mutations if not explicitly controlled during CRISPR-mediated editing. Another pathway which repairs the broken ends of the DNA by ligating it is **MMEJ**, that bargains on the small homologous sequences near the break site. This could be the potential reason for larger deletions or genomic rearrangements mediated by MMEJ in comparison to small indels via NHEJ (Troung et al., 2013; Wang et al., 2017; Cheng et al., 2011; Mateos et al., 2015; Wang et al., 2006; Sfeir et al., 2015; Nakade et al., 2014; Sakuma et al., 2016; Lee et al., 2018). Less active pathway in many cell types is **HDR** which utilizes the exogenous DNA template having homology with the target strand possessing DSBs to repair it precisely. Despite enabling desired gene edits due to the less error-prone nature of HDR, it is still the area of concern among researchers due to its limited genome editing efficiency (Symington et al., 2011; Huertas et al., 2009; Garcia et al., 2011; Daley et al., 2017; Renkawitz et al., 2014; Bhat et al., 2018; San et al., 2008; West et al., 2009; Heyer et al., 2010). **SSA** also uses homologous sequences but usually involves larger deletions between repeated DNA sequences. It's less frequently employed in CRISPR-based editing due to its tendency to cause substantial DNA loss. Directing the CRISPR mediated accurate and precise DNA outcomes by orchestrating repair pathways like maximizing HDR while minimizing error-prone pathways like NHEJ or MMEJ is a central focus (Bhargava et al., 2016; Yamaguchi et al., 1998; Stone et al., 2014).

Independently, it has been entrenched that structures of DNA breaks influence the repair pathways. Cas9 nucleases which creates blunt DSBs preferentially repaired by classical NHEJ and leads to random insertions or deletions. Whilst, 5' staggered bases (More than 30 bases) produced by nickase Cas9 targeting two different positions, promotes resection of DNA, alternate NHEJ and HDR. Engineered FnCas9 creates 5' overhangs ranging from 2-3 bases and also get repaired by HDR

factors; this holds a promise in governing HDR based precise edits.

We concomitantly measured the precise editing and indel frequency of engineered FnCas9 and canonical SpCas9 in combination with Donor templates (varying lengths and types) introducing small sequences at specific loci (*HBB*, *FASN* and *LMNA*) in HEK293T cells. These Cas9 induced alterations were analysed by amplicon sequencing that has quantified precise editing outcomes and indel frequency. To clarify cell line dependence, we also assessed the precise editing outcomes in stem cells considering their longer S phase of cell cycle.

Precise DNA interrogation property of engineered *FnCas9* variants with efficient editing accentuates its efficacy in safe and effectual therapeutic gene correction across multiple cell lines and target loci. We performed *en1 FnCas9* mediated editing in mouse zygotes and in human embryonic stem cells at specific DNA loci like *OCT4*.

## 4.2 Methods and Materials

### 4.2.1 Plasmid Construction

PX408 (Addgene 68705) was used as a template for PCR amplification of the gene encoding the full length *Francisella novicida* (FnCas9 nuclease residues 1–1629 bp) as well as its modified variations. Following that, these genes were cloned using restriction enzyme-based cloning for the pET28-His-10-Smt3 vector (a gracious gift from Prof. Stewart Shuman and Dr. K.M. Sinha) and ligation-independent cloning (LIC) for the pET-His6-GFP-TEV-LIC vector (Addgene 29663), respectively. With minor changes to the manufacturer's procedure, Quick-change II site directed mutagenesis kit (Agilent) was used to create catalytically inactive FnCas9 double mutants on pET His6-FnGFP-TEV-LIC plasmid backbone.

### 4.2.2 Protein purification

As stated earlier, the proteins employed in this investigation were purified. To put it briefly, *Escherichia coli Rosetta2 (DE3)* (Novagen) was used for creating plasmids encoding distinct Cas9 proteins. Until the OD600 reached 0.6, the *Rosetta2 (DE3)* cells were grown in LB media supplemented with 50 mg/l kanamycin at 37°C. The addition of 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) was used to stimulate protein expression. After an overnight cultivation at 18°C, the cells were separated by centrifugation. Re-suspended *E. Coli* cells were lysed by sonication and centrifugation in lysis buffer (20

mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol), supplemented with 1X protease inhibitor cocktail (PIC, Roche), and 100 µg/ml lysozyme.

The lysate was affinity-purified by Ni-NTA beads from Roche, and the eluted protein was further purified using chromatography based on size-exclusion on a HiLoad Superdex 200 (16/60 column) from GE Healthcare in 150 mM KCl, 20 mM HEPES pH 7.5, 1 mM DTT and 10% glycerol.

Thermo Fisher Scientific's Pierce BCA protein assay kit was used to quantify the concentration of purified proteins. Up until their next usage, the purified proteins were kept in storage at -80°C. For 6XHis-MBP-dSpCas9, dFnCas9, and its engineered variants the 6XHis-MBP was eliminated by soaking the affinity-bound protein in cleavage buffer (50 mM Tris-Cl, pH8, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT) for a whole night in order to incubate the protein with PreScission Protease. On a HiLoad Superdex 200 16/60 column (GE Healthcare) with 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, and 1 mM DTT, the cleaved Cas9 protein was isolated from the fusion tag.

#### **4.2.3 *In- vitro* transcription**

Using the Mega Script T7 kit (Thermo Fisher Scientific) and T7 promoters with templates as substrates, sgRNAs were transcribed in vitro. IVT reactions were purified using a NucAway spin column (Thermo Fisher Scientific) after being incubated for a full night at 37°C. Until they were needed again, IVT sgRNAs were kept at -20°C.

#### **4.2.4 *In-vitro* Cleavage (IVC) assay**

For the DNA repair study, guide RNA targeting the survival marker like *ampicillin* and reporter genes like *Zsgreen* have been used in case of bacterial and mammalian cells respectively. Plasmids containing the target sequence (*ampicillin* and *Zsgreen*) and the respective PAM was used as the substrate for *in vitro* cleavage experiments. Plasmids (50 ng or ~5 nM) were incubated at 37°C for 30 mins with the Cas9–sgRNA complex (50 nM) in 10 µL of reaction buffer, containing 20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, and 5% glycerol. The reaction was stopped by the addition of a quenching buffer, containing EDTA (20 mM final) and Proteinase K (40 ng). The reaction products were resolved in EtBr-stained 1% agarose gel, visualized by Syngene UV transilluminator and quantified by ImageJ

#### **4.2.5 Construction of the Cas9 stable cell lines**

To investigate the repair proteome associated with Cas9 following cleavage of double-stranded endogenous target DNA, it is essential to ensure adequate expression of Cas9. This allows for the capture of repair proteins that interact with Cas9. To achieve this, we designed a strategy to generate stable Cas9-expressing cell lines. Using the Addgene parent plasmid pHIV-dTomato (21374), we cloned *SpCas9*, *FnCas9*, and *engineered FnCas9 (e1FnCas9)* into the plasmid using the NotI and EcoRI multiple cloning sites (MCS). After confirming all constructs through Sanger sequencing, transient expression of Cas9 in HEK293T cells was validated by Western blot analysis. Subsequently, the lenti-Cas9-tdTomato constructs were used to produce viral particles in HEK293T packaging cells with envelope and packaging plasmids. These viral particles were then transduced into HEK293T cells to establish stable cell lines expressing SpCas9, FnCas9, and e1FnCas9. Cas9 expression in the HEK293T stable cell lines was quantified using qRT-PCR and Western blot.

#### **4.2.6 Real Time PCR**

Using the Qiagen RNeasy kit (74106) and the comprehensive instructions included in the kit, total RNA was extracted from stable cells expressing Cas9. After preparing cDNA synthesis using the Qiagen kit (205313) for cDNA and following the recommended conditions, the extracted RNA was incubated with DNase I. Real-time qPCR was performed in triplicates on the test and control samples subsequent to cDNA synthesis. TB Green Premium Ex Taq II was used in the PCR apparatus Light Cycler 480 (Roche) or Bio-Rad to quantify the transcripts. Every Ct value acquired for various transcripts was normalized using the GAPDH Ct value. Fold change analysis for comparative examination of the transcripts was done using the 2-DDCt method.

#### **4.2.7 Western blot**

To prepare the cell protein lysate, Cas9 stable cells were lysed using RIPA lysis buffer supplemented with a protease inhibitor cocktail (PIC, Roche cOComplete). A total of 100 µL of lysis buffer was added to each well of a 6-well plate. The cells were incubated on a rocker at 4°C for one hour. Protein lysates were then collected from each sample, and protein concentrations were determined using the Pierce BCA Protein Assay Kit (ThermoFisher).

For SDS-PAGE, 50 µg of protein from each sample was loaded onto an 8% SDS gel. The electrophoresis was conducted using SDS running buffer containing 2.5 mM Tris base, 19 mM glycine,

and 0.1% SDS in autoclaved Milli-Q water.

The proteins were subsequently transferred utilizing the Bio-Rad vertical gel transfer apparatus from the gel to the PVDF membrane using Transfer buffer (2.5 mM Tris base, 19 mM Glycine, 20% v/v Methanol in autoclaved milliQ) at 4°C for 3 hrs at 75 V. Following the completion of the transfer, the membrane was sliced to the appropriate protein size and blocked with 5% BSA in PBST (20 mM Tris base, 150 mM NaCl, and 0.2% Tween 20) at room temperature for two hours at the rocking platform. Following blocking, the blots were incubated for a further night at 4°C in a rocker with the primary antibody (Abcam 210752) at a 1:2500 dilution in the same blocking buffer. The loading control is a 1:1000 dilution of the vinculin antibody. Following the main antibody incubation, the blots were cleaned three times with 0.2% PBST for ten mins each. The identical blots were washed again, and then they were incubated for two hours at room temperature in a rocker with a secondary antibody that had an HRP conjugate. Upon the secondary antibody incubation, the blots had three 15-mins washes. The blots of the Syngene Gel doc instrument were developed using chemiluminescent HRP Substrate (ECL) for signal generation. The densitometry analysis and signal quantification were performed using ImageJ.

#### **4.2.8 Viability and apoptotic assays for Cas9 stable cell lines**

Alamar Blue assay: The human epithelial kidney cell line HEK293T cells were maintained in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin, in an incubator at 37°C, 5% CO<sub>2</sub>; the day before exposure,  $5 \times 10^3$  cells/well were seeded in a 96 well plate. A negative control (NC) and a positive control (PC) have also been included, i.e., cells unexposed to the test substance (maintained in the cell culture medium) and cells exposed to a known cytotoxic agent, which was 30% DMSO in our case respectively. Cells were washed with PBS twice and 10 ul of the AB with 100 ul of Media is added per well after mixing it thoroughly. After adding the AB, the plate was incubated for 1–4 hrs (4hrs in our case) at 37°C, 5% CO<sub>2</sub>, until a change in AB colour was observed. Cell cytotoxicity, its proliferation and metabolic activity was determined by AB assay at 570 nm of excitation after 4 hrs of incubation.

Annexin PI staining: After preparing the 1X binding buffer by mixing 1 part of the 10X binding buffer with 9 parts of distilled water, the cells were harvested. Washing of the cells with 1X binding buffer was done followed by 1X PBS. Cells were resuspended in 1X binding buffer at  $1-5 \times 10^6$  cells/mL. A negative control (NC) and a positive control (PC) have also been included, i.e., cells unexposed to the test substance (maintained in the cell culture medium) and cells exposed to a known apoptotic agent, (4 hours of 5 nM staurosporine treatment) in our case respectively. 5 ul of fluorochrome-conjugated

Annexin V was added to 500  $\mu$ L of the cell suspension. This mix was incubated at room temperature for 15 mins before acquisition in BD LSR II.

#### **4.2.9 $\gamma$ -H2AX staining**

Cells were harvested and resuspended in culture medium to  $5 \times 10^5$  cell/mL. Cells were seeded in each well and incubated overnight at 37°C and 5% CO<sub>2</sub> (cell should be at >80% confluent). Media was aspirated and 500 ul of 1X PBS was added for DNA DSB induction with 30 mJoules of the UVB as a positive control. After 5 mins of incubation, fresh media was added in wells. After 18 hours of UVB treatment, cells were fixed with 4% PFA for 15 mins at RT. Washing was done twice with a washing buffer (30 nM Glycine in PBS, 5 mM EGTA and 10 mM MgCl<sub>2</sub>) and permeabilized with 0.2% Triton X at room temperature for 10 mins. Blocking of the cells were done by 0.5% BSA for 30 mins at room temperature followed by 2X washings. After removing the blocking, cells were incubated with the  $\gamma$ H2A antibody (ab26350) in 1:200 dilution overnight in cold. After 3 successive PBS washings, cells were incubated with secondary antibody (Goat anti-mouse Alexa Fluor 488 with 4 ug/ml) for 1 hour of incubation at RT. After three successive washings with 1X PBS solution was removed and DAPI diluted in PBS was added.

#### **4.2.10 Electroporation in HEK293T cells**

HEK293T cells were cultured in DMEM with GlutaMAX TM supplement (ThermoFisher Scientific Cat. No. 10566016) with 10% FBS serum. 70%-80% confluent HEK293T cells were harvested from a 6 well plate using Trypsin-EDTA (0.05%) (ThermoFisher Scientific Cat. No.: 25300062) and pipetted to make single cell suspension. For each electroporation reaction, total 1 ug Cas9-gRNA vector plasmid with 50 pmoles of ssODN, 25 pmoles dsODN HDR donor template or (1 ug of only gRNA plasmid and 500 ng of linearised donor having GFP in case of Cas9 stables) was mixed in Resuspension buffer R. dsODN donor template was prepared by making a duplex using the Forward and Reverse primers complementary with each other of same lengths via cooling down followed by its denaturation.  $2 \times 10^5$  cells were resuspended in 10  $\mu$ L of Resuspension Buffer R containing DNA and electroporation was performed using Neon® Transfection System 10 ul Kit (ThermoFisher Scientific Cat. No. MPK10096) with single pulses at 950 V, 30 milliseconds pulse width. The electroporated cells were transferred immediately to a 12 well plate containing 1 ml of pre-warmed culture medium and incubated at 37°C and 5% CO<sub>2</sub>. After 24 hours cells were washed and re-incubated with fresh culture medium. 72 hours post electroporation GFP positive cells per sample were quantified in case of HDR analysis and sorted

using BD FACSMelody Cell Sorter (BD Biosciences-US) and gDNA was isolated from the sorted cells using Lucigen buffer for NGS sequencing.

#### **4.2.11 Culturing of hESCs**

hESCs cells were cultured on Matrigel, fed daily with mTeSR1 (minimal Essential Medium supplemented with Thrombopoietin, Epidermal Growth Factor, and other factors), passaged every 3-4 days using REleSR (Reprogramming and Long-term Expansion of Stem Cells in Reproducible, Efficient Stem Cell Medium) at a 1:10 - 1:20 ratio. For Plating these cells enough mTeSR1 supplemented with CloneR2 (10x) was prepared and equilibrated to room temperature. After aspirating media from wells, cells were washed with PBS. Following this accumax was added per well and incubated at 37°C for 15 mins to weaken cell clumps. After checking the cells became loose, accumax was quenched by adding 1 ml mTeSR1/CR2 to each well. Cells were disaggregated to a single cell density by gently pipetting up and down. Very few/ virtually no clumps were there and cells were transferred into a fresh tube with 5-10 ml of mTeSR1/CR2 inside to aid survivability.  $3 \times 10^5$  cells for each well were plated in 1 ml per well of a pre-coated Matrigel 12 well plate and placed at 37°C. After 18-24 hours of cell inspection, Colonies of 4-10 cells were emerging. After ensuring that colonies were not too large and no obvious clump-derived colonies were present, cells were abundant and forming a good ‘web’ structure transfection was performed following this.

#### **4.2.12 Transfection of Cas9 editing components using lipofectamine**

Cas9 plasmid DNA containing guide RNA (750 ng) was combined with 50 pmol of single-stranded oligonucleotide (ssODN) and 25 pmol of double-stranded oligonucleotide (dsODN) HDR templates. The mixture was prepared using Lipofectamine Stem reagent and Opti-MEM, followed by incubation at room temperature for 10 mins. Fresh mTeSR1 medium was added to each well, and the transfection mixture was introduced dropwise.

Cells were incubated for 30 hours post-transfection and then subjected to selection with 0.5 µg/mL puromycin for 2 days. After 48 hours of puromycin treatment, fresh mTeSR1 medium was added, and the cells were cultured until they reached 60–70% confluency. Cells were collected on the same days to ensure no temporal advantage between samples.

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit, and the OCT4 locus was amplified

via targeted PCR. Amplicons were purified using AMPure XP beads. Equal volumes of beads were added to the samples, mixed thoroughly, and incubated at room temperature for 5 mins. The samples were placed on a magnetic stand for 2 mins to clear the supernatant, and the beads were washed three times with 80% ethanol. After air drying, the purified amplicons were eluted in 10 mM Tris buffer (pH 8.5, EB buffer).

#### **4.2.13 Immunostaining of Human stem cells**

Cells were rinsed with PBS and fixed with 4% paraformaldehyde (PFA) for 1 hour at 4°C. After fixation cells were washed quickly three times in 1X PBS + 0.1% Tween. Permeabilization of cells with 1X PBS + 0.5% Tween by incubating the cells for 20 mins on a rotating shaker. Blocking was done with 10% Donkey serum diluted in 1X PBS + 0.1% Tween for 1 hr of incubation on a rotating shaker. Primary antibody incubation was done overnight at 4°C on a rotating shaker (Primary antibody is typically diluted 1:500 in blocking solution). Cells were washed with 1X PBS + 0.1% Tween quickly three times. After permeabilization with 1X PBS + 0.5% Tween secondary antibody was incubated for 1 hr on a rotating shaker (Secondary antibody is diluted 1:300 in 1X PBS + 0.1% Tween). After three successive washings with 1X PBS + 0.1% Tween, solution was removed and Vectashield Mounting Medium with DAPI (diluted 1:30 in 1X PBS + 0.1% Tween) was added.

#### **4.2.14 Cell culture of mESCs**

Mouse embryonic stem cell line, R1/E cultured in feeder-free complete media containing 1X DMEM (Gibco, 11995-065), 20% Fetal Bovine Serum (Pansera Pan Biotech, P29-0705-ES), 1X Pen-strep (Gibco, 15140122), 1% NEAA (Gibco, 11140050), 0.1% BetaMercaptoethanol (Gibco, 21985023) with additional supplement of 8 ng/ml LIF (MPI-CBG, Dresden, Germany) per 500 mL of media. The mESCs were cultured up to a maximum of 20 passages. Media were changed every day and split every alternate day by detaching with 0.25% Trypsin-EDTA (Gibco, 25200056). Cells were incubated in a humidified incubator at 37°C and 5% CO<sub>2</sub>.

Cells were seeded in 12 well plates 2\*10<sup>4</sup> seeding density respectively. 16-20 hrs post seeding, cells were transfected with 1ug per well (12 well plate) of plasmid concentration with the transfection reagent Lipofectamine 3000 (Invitrogen, L3000001) and the protocol mentioned in the kit was followed. 24-48 hrs post transfection, cells were harvested for downstream experiments.

#### **4.2.15 CRISPR/Cas9 editing in mouse zygotes**

sgRNA production and ribonucleoprotein preparation, using the Bbs1 restriction site, the sgRNA was cloned into the bicistronic expression vector px330 (Addgene; 4223037). Using the Q5 hot start high fidelity DNA polymerase (NEB; M0493), the sgRNA sequence from the correctly targeted px330 vector was amplified. The PCR product was subsequently in vitro transcribed using the HiScribe® T7 High Yield RNA Synthesis Kit, and purified using Zymo RNA Clean & Concentrator columns (Zymo Research; R1017). 800 ng/ul of pure Cas9 protein and 400 ng/ul of sgRNA were reconstituted in a tube with Opti-MEM. This was incubated at RT for 15 mins and added with the 100pmoles of ssODN and dsODN HDR templates, pulse spun and ready for electroporation.

Mouse zygote collection- Super-ovulation was performed on four- to eight-week-old (C57BL6×CBA) F1 female mice by injecting five IU of pregnant mare serum gonadotropin (PMSG; Sigma-Aldrich). Five units of human chorionic gonadotropin (HCG; Sigma-Aldrich) were given 48 hours following the PMSG injection. Super ovulated females were paired with F1 males (C57BL6× CBA) who were at least eight weeks old for mating. A 12-hour light-dark cycle was used to sustain the mice. Mouse zygotes under mineral oil (Origio; ART-4008-5P) were isolated in Global total using HEPES (Life Global; LGTH-100), and cumulus cells were eliminated using hyaluronidase (Sigma-Aldrich; H4272). Every animal study project was carried out in accordance with UK Home Office Licence Number 70/8560.

Zona Thinning and electroporation in mouse zygotes- Transferred the necessary number of embryos and washed into FHM media 2-3 times. Transferred embryos into the FHM small drops lid for zona thinning using acidified Tyrode's solution. Washed thoroughly in 9 drops of OptiMEM ahead of the electroporation. Added ~7 ul of the RNP solution having HDR templates to the intra- electrode space on the electrode device Program the Nepa21 electroporator to a voltage of 20 V, pulse length of 25 ms, pulse interval of 50ms and number of pulses to 2. Checked the impedance of the solution to get between 0.20 – 0.24. Pressed Start to trigger a pulse Note down the actual Voltage, Amplitude, Joules and Ohms of the pulse. After this, embryos were taken out from electrodes and washed in 6 drops of FHM media then in 6 drops of Global media, leaving up to 30 embryos per drop for culturing. Pre-equilibrated Global media (Life Global; LGGG-20) was used to culture mouse embryos, and it was topped with mineral oil (Origio; ART-4008-5P) and supplemented with 5 mg of protein per millilitre

(Life Global; LGPS-605). Pre-implantation embryos were cultured for three or four days at 37 °C and 5.5% CO<sub>2</sub> in an Embryoscope+ time-lapse incubator (Vitrolite). Upon collection four days later, the

quantity of blastocysts, morulae, arrests, and lysed was recorded. 25  $\mu$ m of 50 mM NaOH was used for blastocyst lysis, which was then incubated for five mins at 95°C. Samples were vortexed, and 2.5  $\mu$ l of 1M Tris buffer pH 8 was added. Set up the PCR by incubating 5–10  $\mu$ l of the lysate sample at room temperature for the whole night.

#### 4.2.16 Amplicon sequencing

HEK293T cells on six well dishes were transfected with total 1  $\mu$ g Cas9-gRNA vector plasmid with 10 pmoles of ssODN, dsODN HDR donor. 72 hrs post-transfection GFP-positive cells were FACS sorted (BD FACS Melody Cell Sorter) and gDNA was isolated (Lucigen Quick Extract Extraction solution). Amplicons were prepared with site-specific primers and were indexed with Illumina TruSeq i5 and i7 index adapters. Indexed amplicons were purified with AmPure XP beads. Concentrations were determined with Thermo HS-assay kit (Thermo scientific) in a Qubit instrument and amplicons were pooled to make a sequencing library. Pooled sequencing library was sequenced on a Illumina Novaseq 6000 instrument with 2x150 bp chemistry at required depth. Reads were demultiplexed on the instrument. Sequencing reads were quality-checked with FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.14 (<http://dx.doi.org/10.1101/bioinformatics/btw354>). Poor-quality bases with Phred score less than Q30 and TruSeq adapter sequences were removed using cutadapt v2.8 (<http://dx.doi.org/10.14806/ej.17.1.200>) for paired-end reads. Trimmed reads with at least 45 bp length were retained. To check for indels, cleaned reads were analysed with CRISPResso2 v2.2.14 (<https://anaconda.org/bioconda/crispresso2>) with default options, except “-qwc” flag was used to restrict the quantification window to the gRNA region and “--ignore\_substitutions”. CRISPR-DAV v2.3.4 (<https://doi.org/10.1101/bioinformatics/btx518>) was used for indel and HDR quantification with default options. An in-house Python script was used to combine editing outcomes from different samples into a CSV file. Data analysis was done in R v4.3.2 and plotted using ggpubr v0.6.0.

Quantification of insertion types- For each target site, cleaned paired-end reads were analysed for indels with CRISPResso2 with above mentioned options and resulting “Alleles\_frequency\_table\_around\_sgRNA\_[sgRNA\_seq].txt” file was used for downstream analysis. A custom in-house Python script was used to quantify what base was inserted at the cleavage site during DNA-repair and the associated frequency. Briefly, it compares the wild-type allele against edited alleles with insertion and looks for inserted bases by alignment. Successively it outputs all the different types of insertions and their frequency.

#### **4.2.17 Long range nanopore sequencing and analysis of large deletions**

Amplicon generation: Specific primers targeting the HBB region and containing the guide RNA site were designed and amplified 5kb region. The amplicon-based sequencing was performed using the Ligation Sequencing Kit (SQK-LSK109) from Oxford Nanopore Technology (ONT). Individual amplified fragments were quantified using Qubit 1X dsDNA HS Assay kit (Thermo Fisher Scientific, Cat. No. Q33230). An input of 100 ng was then purified and size-selected using 1.8X AMPure XP beads (Beckman Coulter, USA). The subsequent library preparation steps were performed following the Native barcoding of amplicons as per manufacturer's recommendation (native-barcoding-amplicons-NBA\_9093\_v109\_revO\_12Nov2019). Ultra II End-prep reaction buffer and Ultra II End-prep enzyme mix were added to the samples as specified, followed by mixing and centrifugation to repair and end-prep the DNA. Following cleanup using 1.2X AMPure XP beads, the end-prep product underwent native barcode ligation (EXP-NBD114). Afterward, barcoded products from the samples were pooled and subsequently purified using AMPure XP beads. The resulting purified library was ligated with an adapter protein and loaded on the R9.4.1 flow cell and sequenced using MinION Mk1C (ONT) platform with a loading concentration of 100 ng of the library.

Detection of structural variation (SV)- The raw fast5 files from ONT MinION were base called and demultiplexed using Guppy base caller that uses the base calling algorithms of Oxford Nanopore Technologies (Nanopore Community) with phred quality cut-off score >7 on GPU-linux accelerated computing machine. Reads with a Phred quality score of less than 7 were discarded to filter the low-quality reads. The resultant demultiplexed fastq were used for further SVs detection and aligned to the reference sequence of targeted HBB gene locus using the aligner Minimap2 v2.26 (Li, 2018). Variant detection with mapped libraries was performed on each individual sample using CuteSV (v2.0.3) (<https://github.com/tjiangHIT/cuteSV>) (Jiang et al., 2020) with the following parameters: –max\_cluster\_bias\_INS 100 --diff\_ratio\_merging\_INS 0.3 --max\_cluster\_bias\_DEL 100 –diff\_ratio\_merging\_DEL 0.3 --mins\_mapq 7 --report\_readid --mins\_support 5 --mins\_size 5 –max\_size 100000 --genotype. Samtools v1.19 (Danecek et al., 2021) was employed to convert, sort, and index bam files. The vcf files from the CuteSV output were used to get the variant Positions. BAM files of aligned Nanopore reads were visually inspected using IGV Browser v2.17.1 for large deletions and for plots (Li et al., 2018; Jiang et al., 2020; Danecek et al., 2021).

#### **4.2.18 Digenome sequencing in mESCs**

Genomic DNA (gDNA) from mouse embryonic stem cells was purified using QI Amp DNA mini kit (Qiagen Cat No. 51306). 3 µg gDNA was incubated with either 500 nM respective Cas9 RNPs (Cas9: sgRNA, 1:2) or a mock sample containing nuclease free water in 1x rCutSmart Buffer (NEB) and 5% glycerol in 500 µL reaction volume at 37°C for 10 hours. gDNA was purified using QI Amp DNA mini kit (Qiagen Cat No. 51306) after quenching the reactions with RNase A (Ambion, 10 mg/mL) and Proteinase K (Invitrogen, 20 mg/mL). The purified gDNA was used for quantitative PCR using Light Cycler 480 SYBR Green I Master (Roche) to confirm the on-target cleavage efficiencies w.r.t the mock sample following which the gDNA of both the Cas9 treated and mock samples were sheared using S220 Focused- ultrasonicator (Covaris) with a median size of the DNA fragments around 250 to 300 bp (Fill 10, Duty 10, PWP 140, CPB 200, 120 sec, water temperature 4°C). DNA library was prepared by NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB #E7645) using the fragmented DNA essentially following the manufacturer's protocol with some modifications. Libraries were indexed by NEBNext® Multiplex Oligos for Illumina® (NEB#E7600) following limited cycle PCR as per manufacturer's protocol. Indexed libraries were cleaned up using AMPure XP beads (A63881, Beckman Coulter). The libraries were quantified either by Qubit dsDNA HS Assay kit (Invitrogen, Q32853) or NEBNext® Library Quant Kit for Illumina® (NEB E7630) and pooled. The qPCR cycling conditions on the instrument, Light Cycler® 480 System (Roche) were as follows: Initial denaturation 95°C for 1 mins followed by 35 amplification cycles of 95°C for 15 sec; 63°C for 45 sec and melt curve. 2× 150-bp sequencing was performed on Illumina NovaSeq 6000 platform at 30-40x depth at Life Cell Diagnostics (Chennai, India).

Digenome-seq analysis- Sequencing was done on Illumina NovaSeq 6000 platform with 2x150bp paired-end sequencing chemistry. Sequencing reads were quality-checked with FastQC v0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and MultiQC v1.14 (<http://dx.doi.org/10.1093/bioinformatics/btw354>). Poor-quality bases and TruSeq adapter sequences were removed using cutadapt v2.8 (<http://dx.doi.org/10.14806/ej.17.1.200>) for paired-end reads.

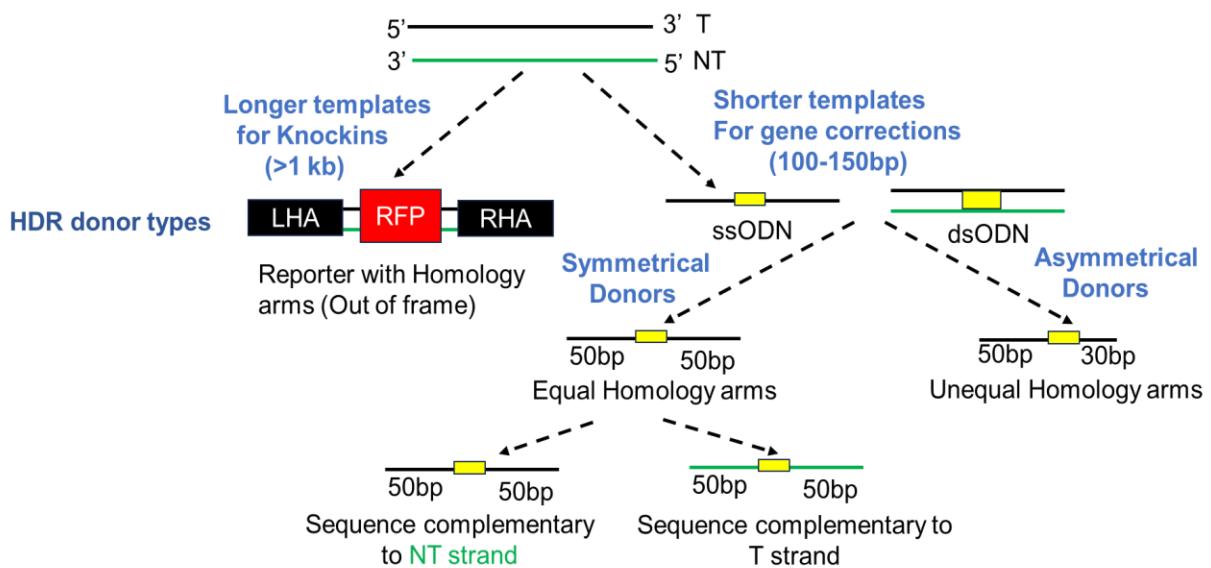
Trimmed reads with at least 45 bp length were retained. Processed paired-end reads were aligned to mm10 genome with bowtie2 v2.2.5 (<https://doi.org/10.1038/nmeth.1923>) with default options and position-sorted with samtools v1.18 (<https://doi.org/10.1093/bioinformatics/btp352>). Aligned reads to only canonical chromosomes were used for all downstream analysis. A command line version of the

Digenome-seq web tool (<https://doi.org/10.1038/nmeth.4262>) was used for analysis with default options except, for SpCas9 the overhang value of 0 and for en1FnCas9 enzyme 3 was used. Cut-off for Digenome-seq score was manually checked from depth of aligned reads for all the cas9 enzymes. Hits from Digenome-seq were further filtered with an in-house Python script to exclude possible false positives.

## 4.3 Results

### 4.3.1 Stably expressing en1FnCas9 mediated staggered cleavage leads to higher HDR knockin rates at LMNA and FASN locus

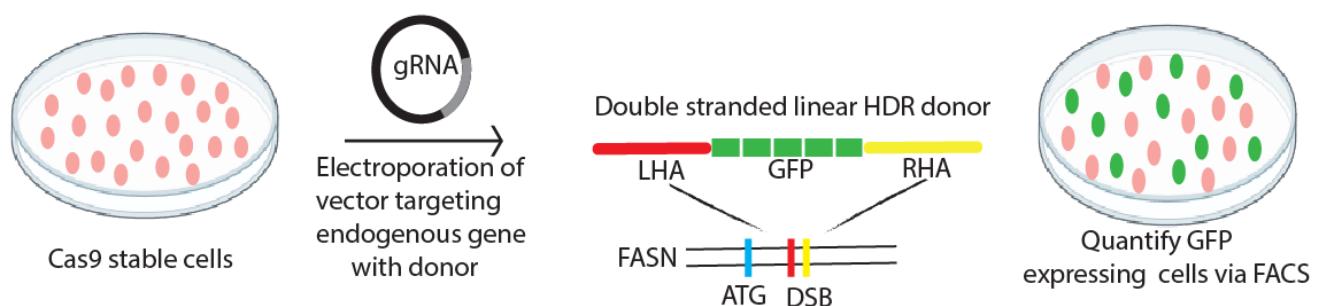
Precise genome editing embarks with the DSBs generation at the targeted site where the specific DNA edit is desired and it has been hastened by the invention of CRISPR-Cas9 mediated programmed DNA cleavage by a guide RNA (crRNA). DSBs can be repaired by Homology Directed repair utilizing the donor DNA molecules flanked by homologous sequences with the target. These DNA donor templates range from short oligonucleotides (ssODNs) of 100-150 bases or large plasmid donors (kbs) depending upon the type of edits. Literature suggested that the HDR efficiency is sensitive to the insert size and this could be because of different engagement of donor types with the target DNA (Yang et al., 2020) (Figure 4.1).

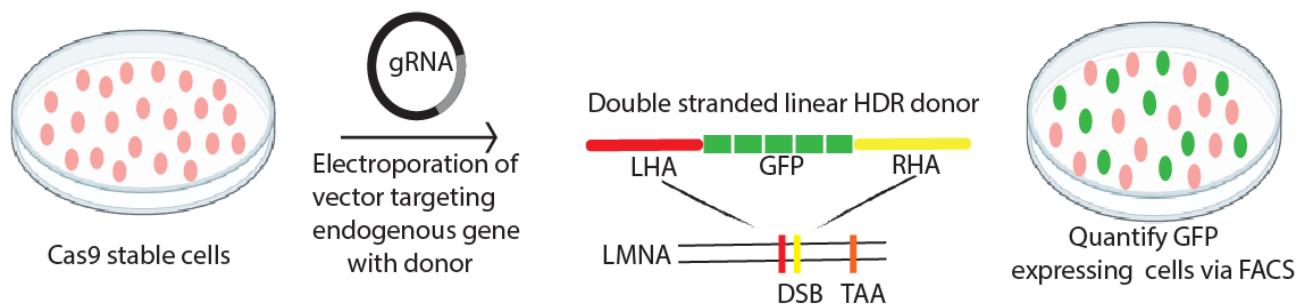


**Figure 4.1 Experimental paradigm showing different lengths and types of donor templates that repair the Cas9 induced DNA breaks via HDR.** Longer templates (>1kb) for knockins or endogenous gene tagging with reporters like RFP, GFP etc. Shorter templates (100-150 base pairs) of single strand ODNs (ssODN) and double strand ODNs (dsODN) for gene corrections with equal (symmetrical) and unequal (asymmetrical) homology arms. Symmetrical donors can have its sequence complementary to either non-target strand of the DNA or target strand for undergoing HDR.

FnCas9 and its engineered variant creates 5' overhangs upon DNA cleavage and we tried to investigate if this Cas9 can improve the HDR in a targeted manner via GFP knockins at endogenous loci and ssODN or dsODN for HDR based therapies. We presumed that the sticky ends generated by en1 FnCas9 might have potential implications in HDR- mediated genomic insertions compared to widely used SpCas9.

Overall, it provides a better insight of en1 FnCas9 mediated efficient and precise genome editing in mammalian cells. We have transfected a linear donor DNA containing an out-of-frame GFP sequence flanked by homology sites to the mammalian FASN and LMNA loci, in stably expressing SpCas9 and en1 FnCas9 HEK293T cells. For producing the FASN and LMNA crRNA in the Cas9 expressing cells we have designed and cloned the crRNA using the BbsI restriction enzyme based cloning strategy. Both the Fn and Sp Scaffold constructs have been sequence confirmed and were introduced in its corresponding Cas9 stable cells. Precise HDR events have been estimated by quantifying GFP positive cells in the FITC channel of BD FACS Melody Cell Sorter (BD Biosciences-US) after 72 hours post electroporation (Figure 4.2).

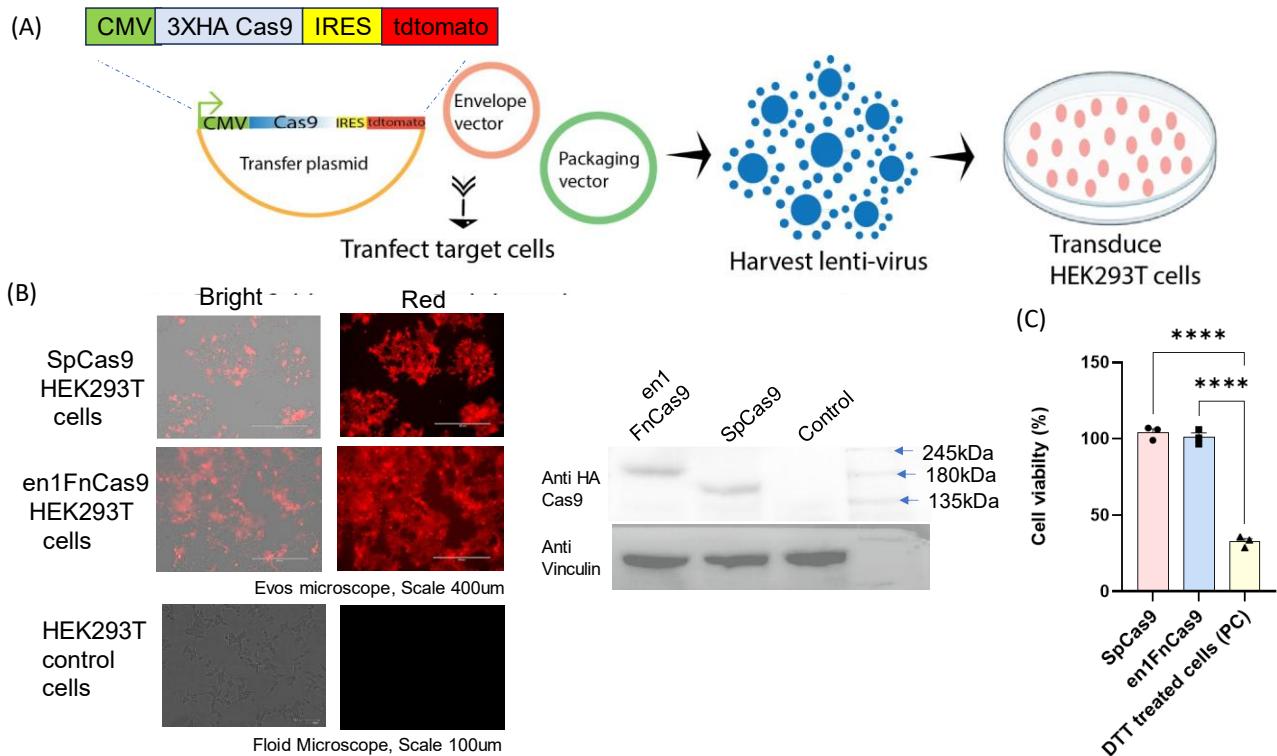




**Figure 4.2 Experimental paradigm showing the Cas9 expressing stable cells undergoing HDR repair pathways in presence of longer DNA template:** Reporter (*mcherry*) positive Cas9 expressing cells electroporated with guide RNA plasmids and linear double stranded donor DNA containing GFP flanked by 250bp homology arms introduced at N terminus of FASN locus in HEK293T cells (Above). Reporter (*mcherry*) positive Cas9 expressing cells electroporated with guide RNA plasmids and linear double stranded donor DNA containing GFP flanked by 250bp homology arms introduced at C terminus of LMNA locus in HEK293T cells (Below). Quantification is done by estimating GFP positive cells in the FITC channel via FACS.

#### 4.3.1.1 Stably expressing SpCas9 and en1 FnCas9 cells are healthy and viable

Few findings suggested that expression of Cas9 protein alone in mammalian cells can activate the DNA damage response because of DSBs induction and lead to cellular consequences like apoptosis and differentiation. This process is sufficient to initiate the  $\gamma$ H2AX foci (Phosphorylated H2AX at serine 139 in its C-terminal tail upon the occurrence of DSBs) formation that is the signature of DSBs in the genome (Xu et al., 2020). It reinforced us for the careful examination of genome stability status in constitutively expressing Cas9 stable cells. We have made the lentivirus based Cas9 expressing HEK293T cells and confirmed its expression with the western blot, observed similar expression of SpCas9 and en1 FnCas9 which reduces the chances of any Cas9 expression biased editing outcomes. We then checked the viability of these cells by Alamar blue assay where we have observed more than 95% viability of Cas9 expressing HEK293T cells and less than 50% DTT treated cells as positive control of the assay (Figure 4.3).

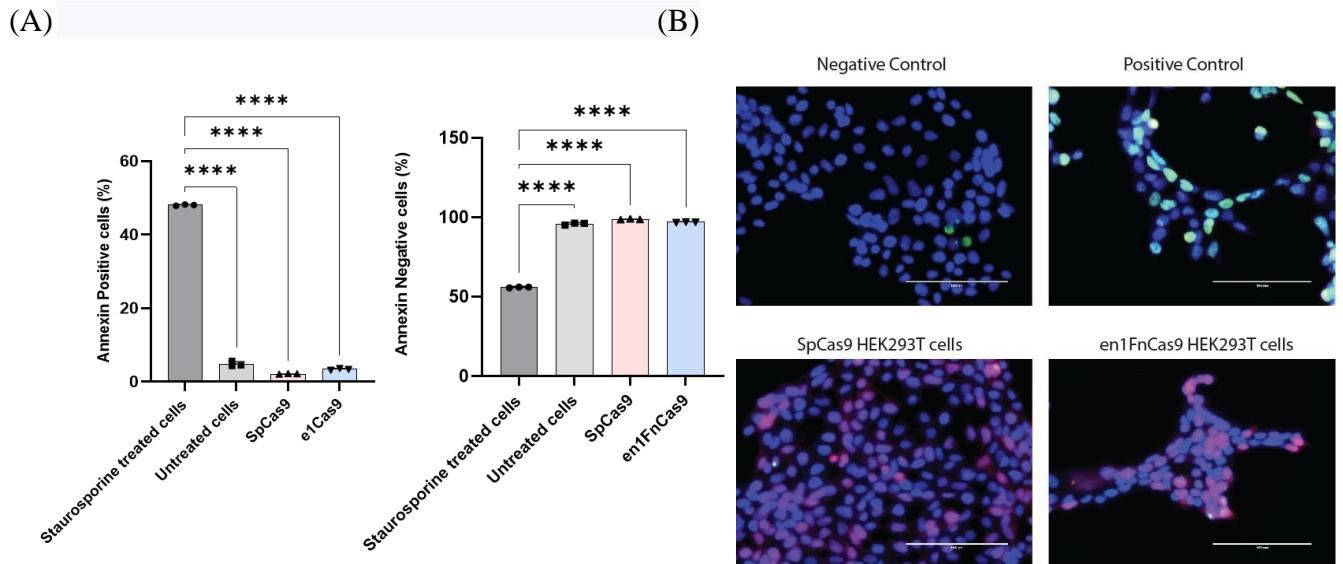


**Figure 4.3 Generation of stable Cas9 expressing HEK293T cells.** (A) Schematic illustration of lentivirus-based overexpression of Cas9 in HEK293T cells. (B) Microscopic images (Left) with scale bar 400 um and Western blot (Right) showing the expression of SpCas9 en1Cas9 in HEK293T cells as 160 kDa and 197 kDa with the control cells respectively. Vinculin taken as a loading control. (C) Bar plots showing the viable percentage stably Cas9 expressing HEK293T cells significantly different from DTT treated cells as a positive control of alamar blue assay. Error bars indicate the SEM of three independent replicates and plotted on GraphPad prism. One-way Anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001.

#### 4.3.1.2 Stably expressing SpCas9 and en1FnCas9 cells don't exhibit visible DNA damage

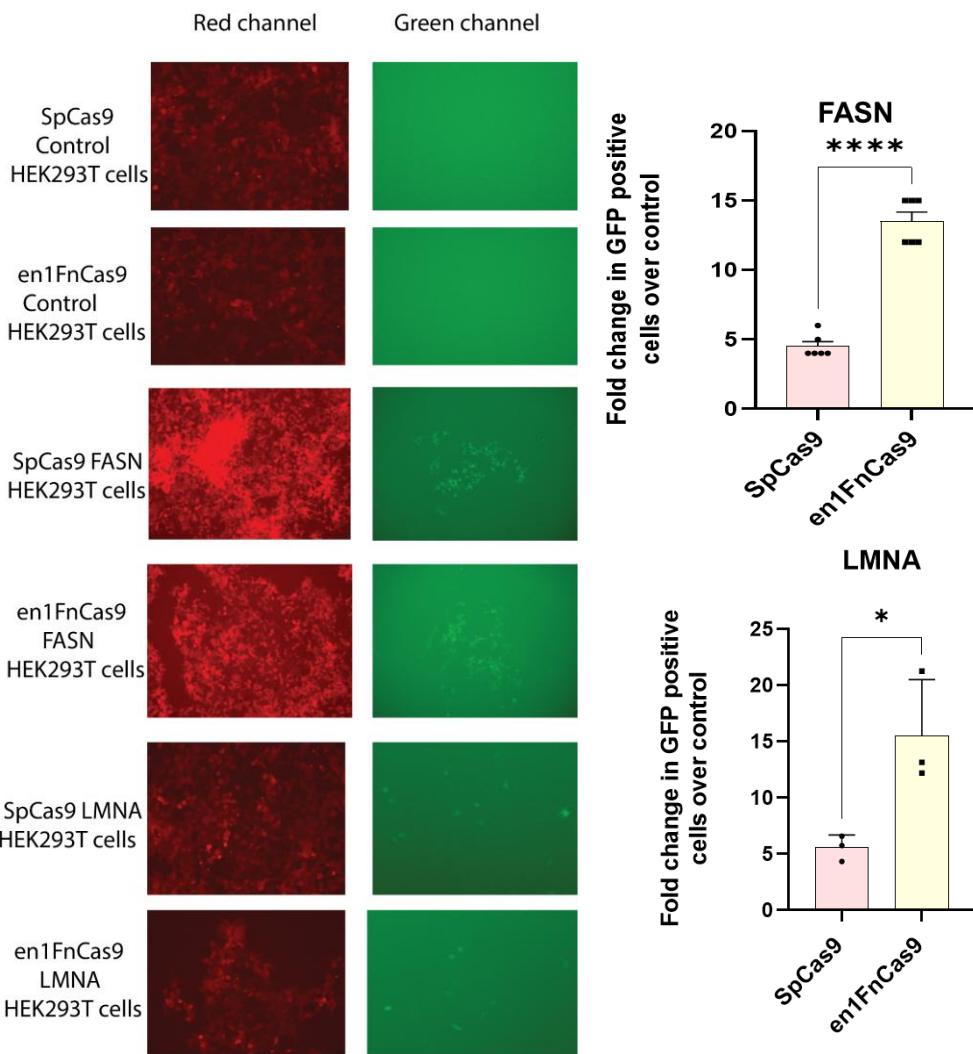
In the field of stem cell biology, Cas9 expression alone orchestrates the phosphorylation of key proteins like p53 (Ser15), CHK1 (Ser317), and H2AX (Ser139) which are the results of DNA damage responses. Activation of *p53*, which is a key regulator of cellular responses to DNA damage, leads to the upregulation of *p53* target genes involved in apoptosis and cell differentiation, such as, *p21*, *NOXA*, *PUMA* and *PERP* (Ihry et al., 2018). After knowing the viability of these cells, we tried to investigate the apoptotic status of these cells by incubating the Cas9 stable cells with the annexin V conjugated with fluorescein isothiocyanate (FITC) which labels the phosphatidylserine sites on the membrane surface of

early apoptotic cells. We have observed negligible percentage of annexin positive cells in case of Cas9 stables apart from the staurosporine treated cells which acted as a positive control in this assay. From the microscopic images of  $\gamma$ H2AX staining of SpCas9 and en1 FnCas9 expressing HEK293T stable cells, we have not seen any  $\gamma$ H2AX puncta hence, no visible DNA damage in these stables (Figure 4.4).



**Figure 4.4 Cas9 expressing stable HEK293T cells are not prone to apoptosis.** (A) Bar plots showing the percentage of Annexin positive (Left) and Annexin negative (Right) cells significantly different from staurosporine treated cells (proapoptotic agent). Error bars indicate the SEM of three independent replicates plotted on GraphPad prism. One-way Anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001. (B) Microscopic images showing the  $\gamma$ H2AX foci (Green) in case of UV treated HEK293T cells (Positive control) and negative in case of SpCas9 and en1FnCas9 mcherry positive stable cells, similar to the untreated HEK293T cells (Negative control). DNA is counterstained with DAPI and represented in blue. Scale bar 50 $\mu$ m.

HDR is the key component of therapeutic genome editing; therefore, we have examined the potential impact of staggered ends by en1 FnCas9 on its efficiency. We have persistently noticed **higher (~3 fold) GFP+ cells in en1FnCas9-transfected cells** compared to well-studied SpCas9-transfected cells targeting independent endogenous loci like *FASN* at N terminus and *LMNA* at C terminus, suggesting higher rate of homology-directed repair (HDR)-mediated genomic insertions (Figure 4.5).

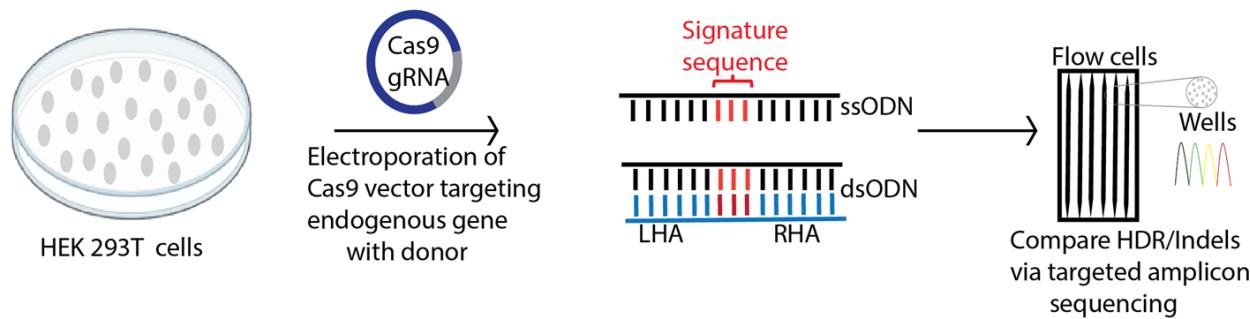


**Figure 4.5 Frequency of HDR is higher during en1FnCas9 mediated editing.** (Left panel) Microscopic images showing the SpCas9 and en1FnCas9 T2A mcherry expressing cells undergoing HDR, subpopulation of these cells is indicating GFP signal after knockin at the desired loci *FASN* and *LMNA*. Bar plots showing the fold change of GFP positive cells normalized over control cells in both *FASN* and *LMNA* (Right Panel). Error bars indicate the SEM of three independent replicates plotted on GraphPad prism. One-way Anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001.

#### 4.3.2 Staggered cutting en1 FnCas9 suppresses NHEJ and makes homologous pathways dominant

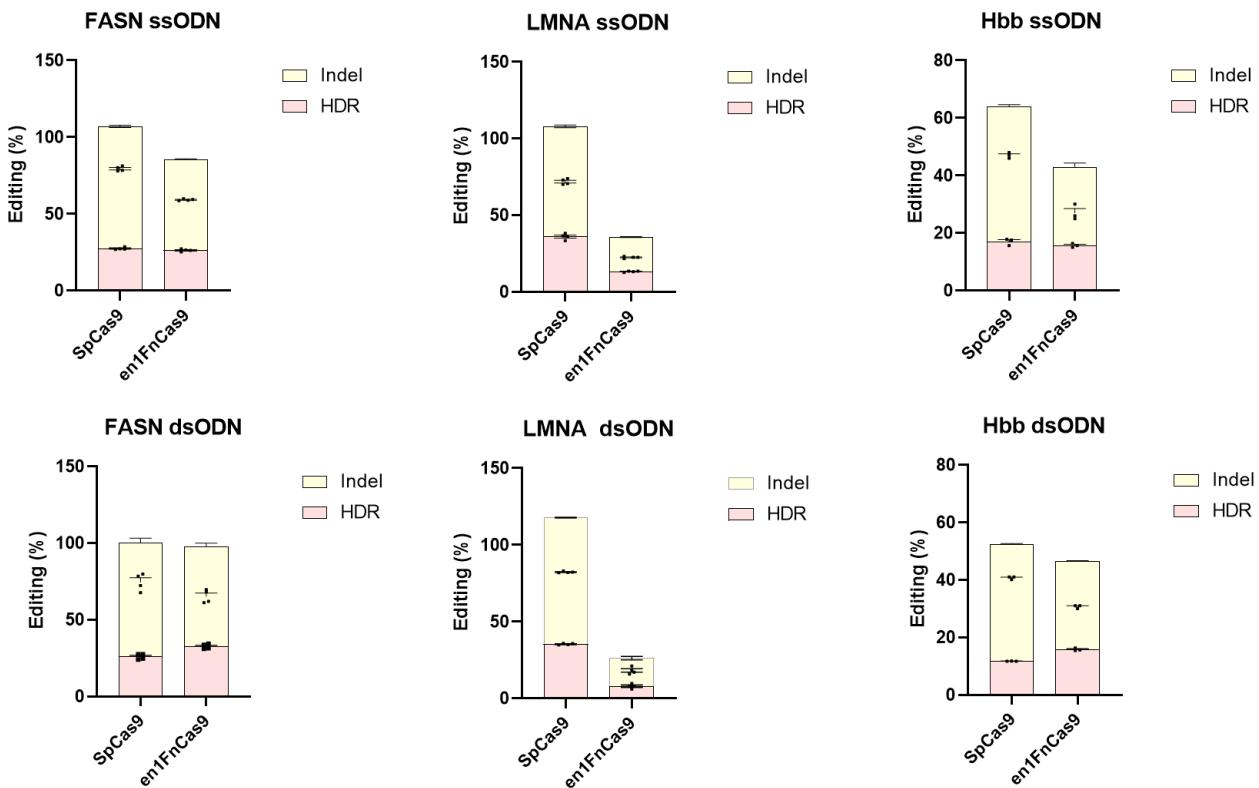
The current focus of CRISPR trials involves ssODN based HDR therapies that introduces short edits and have the possibility of targeting more than 50 percent of known pathogenic mutations according to Clinvar database (Jiang et al., 2022). We next explored the tendency of the staggered ends

preferring the template dependent HDR pathway than the NHEJ in case of short edits based on ssODN. To clarify the mechanisms of these short editing, we subsequently assessed the frequency of HDR via double stranded DNA oligos termed as dsODN of similar length as ssODN. To clarify sequence dependence, we assessed editing at several loci in HEK293T cells using HDR templates producing small edits with analysis by deep sequencing.



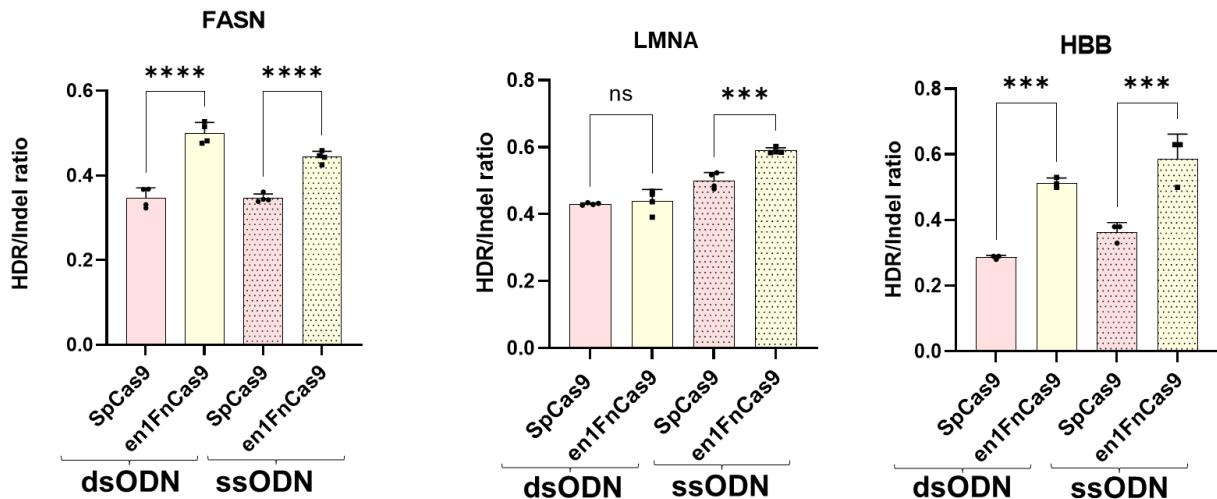
**Figure 4.6 Experimental paradigm of investigating HDR/NHEJ in HEK293T cells.** HEK293T cells electroporated with Cas9 and its guide RNA expressing plasmids, ssODN and dsODN templates are introducing a signature sequence at the precise locus. Frequency of HDR/NHEJ is analysed by deep sequencing.

HEK293T cells were electroporated with Cas9 expressing plasmids fused with GFP, along with the ssODN and dsODN donors of 103bp in lengths having 50 base pairs left and right homology arms with 3 bases of signature sequence like AAA. GFP positive cells that indicated the successful delivery of editing components, were FACS sorted and analysed the editing outcomes at targeted locus using amplicon next generation sequencing (Figure 4.7).



**Figure 4.7 Stacked Bar plots showing the HDR and indel editing percentage in HEK293T cells using ssODN and dsODN templates incorporated at targets (*FASN*, *LMNA*, *HBB*). Types of HDR represented above are perfect HDR events. Error bars represent SEM (3 independent experiments) plotted on GraphPad prism. One-way anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001. (Data Analysed by Prosad Das).**

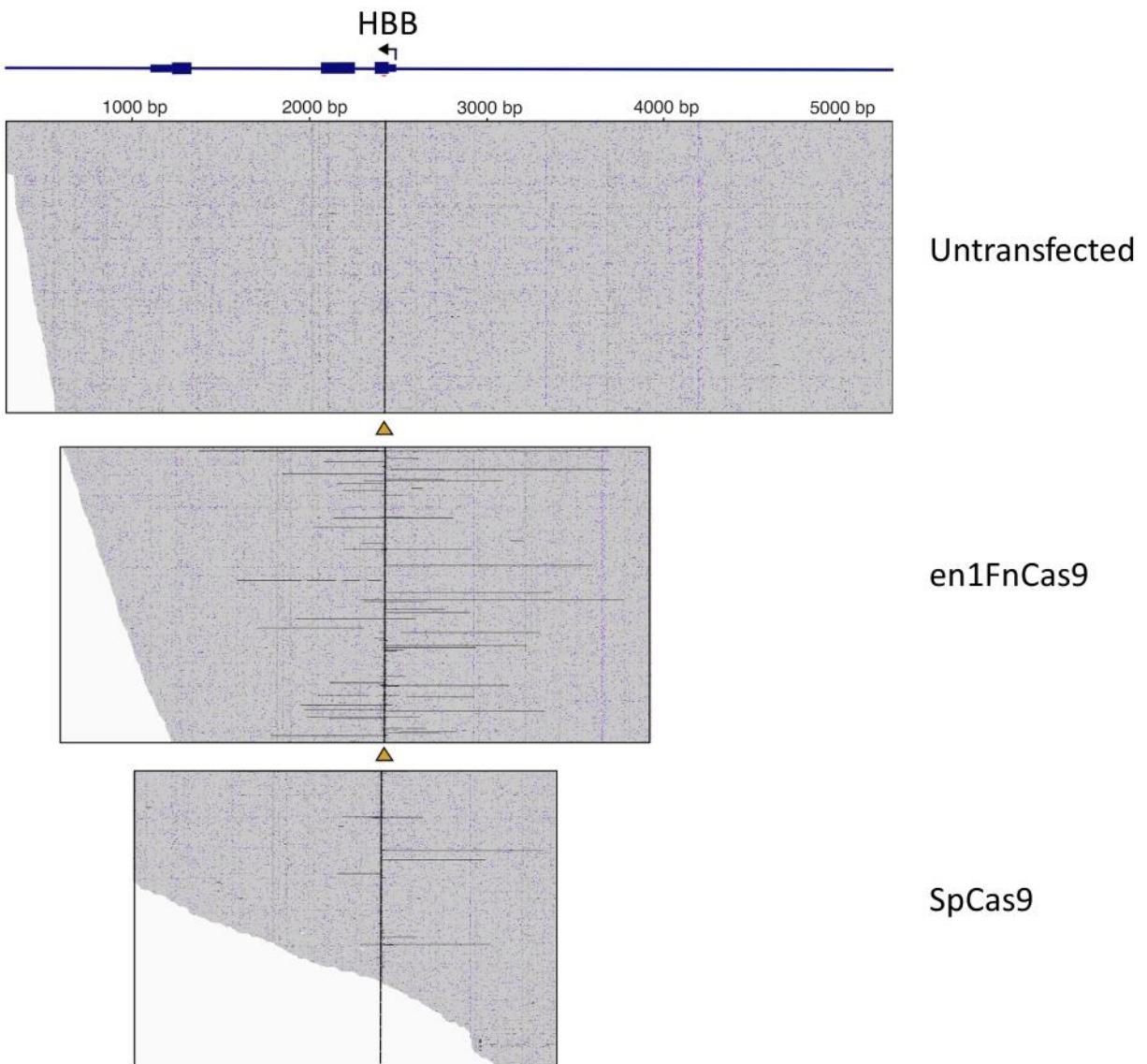
We have observed lesser indels during en1FnCas9 mediated repair in presence of donor templates than SpCas9. The HDR is either comparable (in case of ssODN) or slightly higher (in case of dsODN) during en1FnCas9 mediated editing at *FASN* and *HBB* locus. We did observe drop in indels and HDR in case of en1FnCas9 mediated repair at *LMNA* target site, showing overall less editing at this locus via en1FnCas9 (Figure 4.7).



**Figure 4.8 Bar plots showing the HDR/Indel ratio in HEK293T cells using ssODN and dsODN templates incorporated at targets (*FASN*, *LMNA*, *HBB*). Types of HDR represented above are perfect HDR events. Error bars represent SEM (3 independent experiments) plotted on GraphPad prism. One-way anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001. (Data analysed by Prosad Das).**

We have observed a significantly higher HDR/NHEJ ratio for staggered ended en1FnCas9 than blunt ended SpCas9 cleavage at *FASN*, *LMNA* and *HBB* with the use of both ssODN and dsODN oligos (Figure 4.8).

Despite the distinction of NHEJ and HDR, current studies have divulged the existence of pathways like MMEJ and SSA. Without these homologues' templates, other pathways like MMEJ make use of shorter homologous DNA stretches near the cut site and can induce large deletions compared to c-NHEJ (Gisler et al., 2019). To highlight the intricate interplay between MMEJ and NHEJ pathways in processing the blunt and staggered ended DNA, we have performed the long-range sequencing of the SpCas9 and en1 FnCas9 edited HEK293T cells. We targeted *HBB* locus by SpCas9 and en1FnCas9 in HEK293T cells and amplified the 5000 base pairs of this region, analysed by the nanopore sequencing.



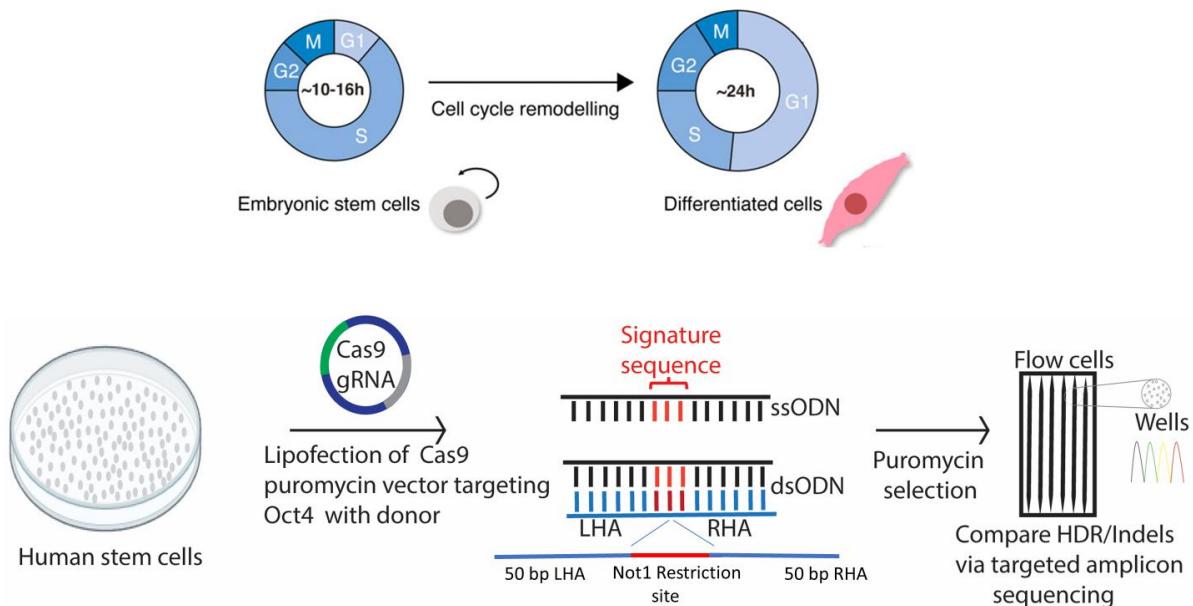
**Figure 4.9 Integrative Genomics Viewer (IGV) screenshot showing the aligned reads to the HBB region (5Kb) detected by nanopore sequencing.** Horizontal black bars represent the deletion lengths in base pairs and vertical dashed lines with the arrow represent the CRISPR Cas9 guide RNA cleavage site. (IGV generated by Prosad Das).

As shown in previous few reports, staggered cutting suppresses small indels caused due to classical NHEJ but makes an alternate pathway prominent which is MMEJ (Owens et al., 2019). We have also observed, as expected, that the number of deletions in the en1FnCas9 targeted *HBB* sample is more than the SpCas9. The un-transfected control amplicons were clean, no deletions have been noticed as depicted by the IGV screenshot. These deletions contained the guide RNA cleavage site (Figure 4.9).

Therefore, it is clear from the study that staggered cutting prefers higher HDR/Indel ratio in presence of donor templates but chooses alternative NHEJ repair pathway (MMEJ) not the c-NHEJ in its absence.

#### 4.3.3 en1FnCas9 mediated HDR editing in human stem cells

Despite the advancements made by CRISPR Cas9, HDR mediated knock-ins still remain a challenge in stem cells due to various factors like cell cycle dependency, donor template design and heterogeneity chromatin Structure. Prevailing over this low cellular efficiency and enhancing the precision, efforts like optimization of experimental conditions, refining the donor template design have been explored (Shin et al., 2020; Lomova et al., 2019). To investigate the staggered ends associated with HDR enrichment strategy in therapeutically more relevant cells like stem cells, we have assessed the targeted repair outcomes in these cells (Figure 4.10).

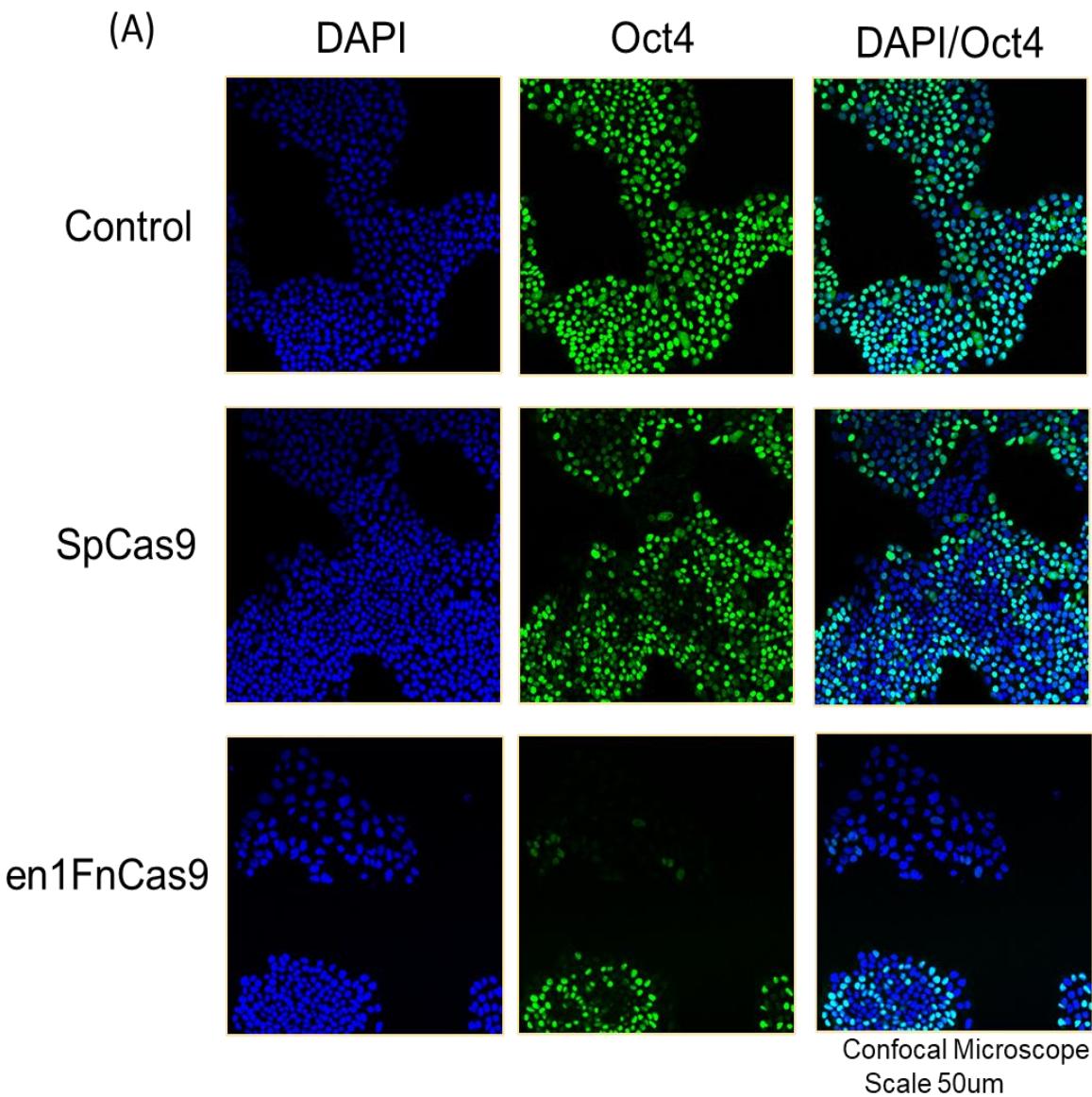


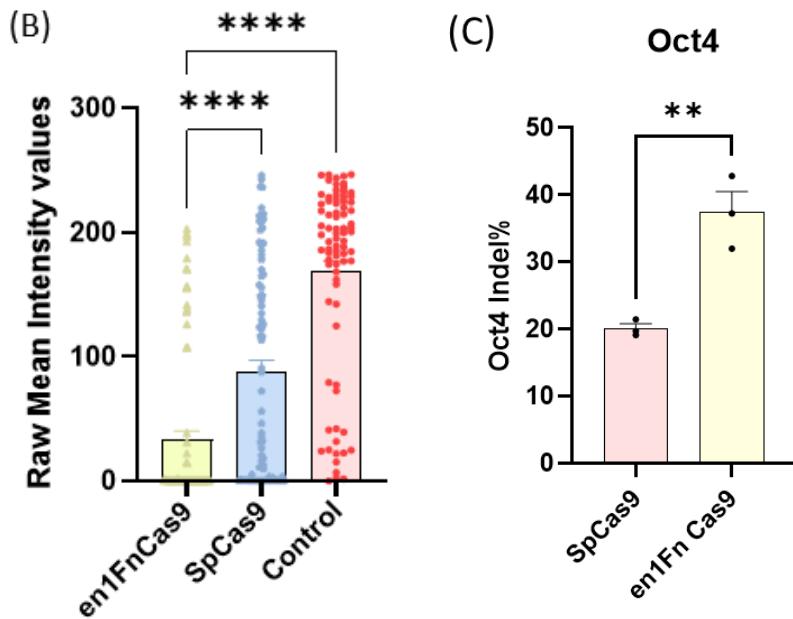
**Figure 4.10 Experimental paradigm of stem cells undergoing HDR in presence of different donor types:** (Upper panel) G1, S, G2 and M-phases of the cell cycle during Stem cell differentiation. (Lower panel) Human stem cells lipofected with Cas9 and its guide RNA expressing plasmids, ssODN and dsODN templates are introducing a signature sequence (Not1 restriction site) at the precise locus. Frequency of HDR/NHEJ is analysed by deep sequencing.

Human embryonic stem cell lines derived from blastocysts (H9) were lipofected with Cas9

expressing plasmids fused with Puromycin, along with the ssODN and dsODN donors of 108 bases in lengths having 50 base pairs left and right homology arms with 8 bases of signature sequence like unique restriction site "GCGGCCGC" (Not1 in our case). Cells upon Cas9 expression became puromycin resistant and were selected for 2 days after 30 hours of lipofection (Figure 4.10).

The repair outcomes at the *Oct4* or *POU5F1* targeted locus were analysed by immunofluorescence for the OCT4 loss of expression and next generation sequencing for absolute indels and HDR events.

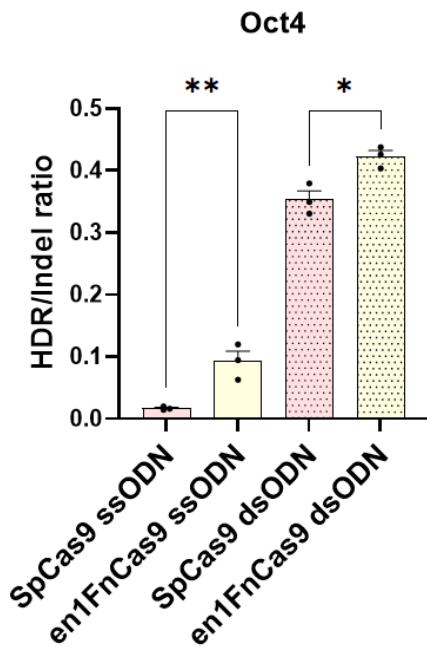




**Figure 4.11 Cellular editing of stem cells by different Cas9 targeting OCT4 locus.** (A) Representative Immunofluorescence images of OCT4 (green) stained human ES cells that are induced to express Cas9 puromycin constructs targeting *Oct4* loci compared to uninduced controls. DNA is counterstained with DAPI and represented in blue. Scale bar 50 $\mu$ m. (B) Bar plots showing the mean intensity values of OCT4 expressing cells N=80, calculated by fiji, in case of Cas9 targeted ESCs. Error bars indicate the SEM of three independent replicates plotted on GraphPad prism. One-way Anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001. © Bar plots showing the indel percentage at *Oct4* loci by different Cas9 analysed by deep sequencing, error bars indicate the SEM of three independent replicates. Unpaired and parametric two tailed t test with p values are shown plotted on GraphPad prism. One-way anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\* p<0.0001. (Data generated in Dr. Kathy Niakan laboratory, University of Cambridge).

We have observed significant loss of OCT4 expression via Immunofluorescence mediated by en1 FnCas9 targeting exon 2 of *Oct4* locus with sgRNA and this loss of OCT4 expression in HESCs depicted the knockout of this gene, hence its failed expression. Deep sequencing data has also shown the higher indel percentage at *Oct4* locus in en1FnCas9 targeted cells than SpCas9 (Figure 4.11).

Eventually, we have experienced similarly significant increase in HDR/NHEJ ratio for staggered ended FnCas9 and en1 FnCas9 than blunt ended SpCas9 cleavage at *Oct4* with the use of both ssODN and dsODN oligos as HEK293T cells (Figure 4.12).



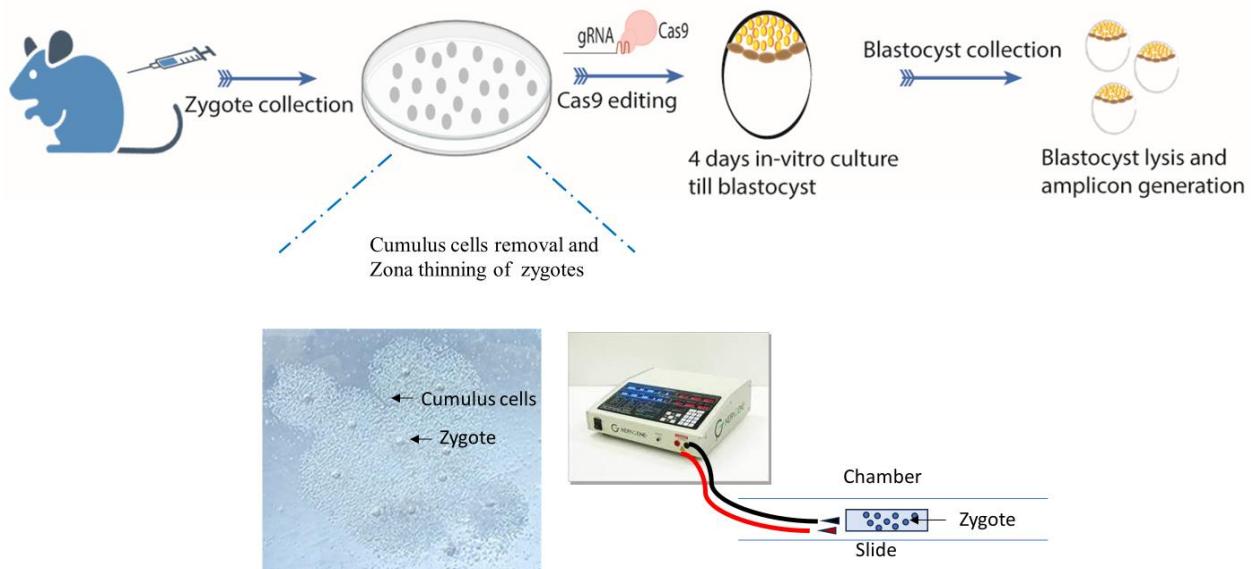
**Figure 4.12 Bar plots showing the HDR/Indel ratio in HESCs cells using ssODN and dsODN templates incorporated at *Oct4* locus.** Pink bar plots showing the HDR/Indel ratio for SpCas9 and yellow bar plots showing the HDR/Indel ratio for en1FnCas9. Types of HDR represented above are perfect HDR events. Error bars represent SEM (3 independent experiments) and plotted on GraphPad prism. One-way Anova P values are shown ns p>0.05, \*p<0.05, \*\*p<0.01.

#### 4.3.4 en1 FnCas9 mediated editing in mouse embryos

Although CRISPR Cas9 is a promising tool for therapeutic genome editing, there are unexpected chromosomal translocations, on target chromothripsis and off-target indels generated that lead to deleterious outcomes and cell death. This is detrimental for its clinical usage (Leibowitz et al., 2021; Cullot et al., 2019). Additionally, to reduce the mutational burden from the progeny, correcting harmful mutations in germ cells or embryos by introducing an efficient and safe Cas9 protein is an attractive avenue. This is currently prohibited by strict regulations since the deleterious and long-term effects of CRISPR based DNA edits have not been completely studied. Embryonic cells have low efficiency of mutation repair and hence a high rate of mosaicism that can lead to unintended editing outcomes. Chromosomal abnormalities and loss-of heterozygosity events in these cells due to Cas9 off site cutting followed by the repair, have recently raised the concern (Fogarty et al., 2017). It is therefore imperative to genome editing pipelines for precise germ line correction and low off-site mutagenesis.

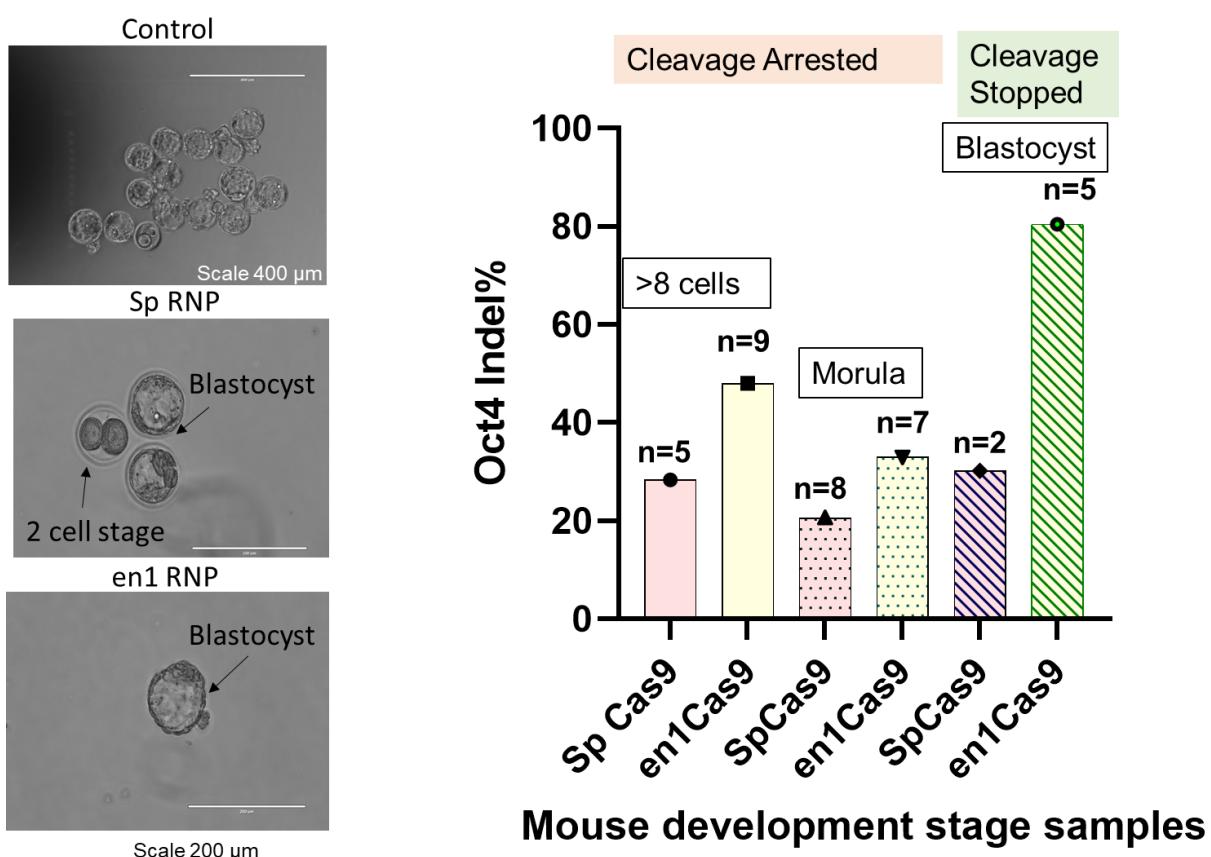
These engineered variants have remarkably shown superior outcomes in terms of knock-in (HDR) rates potentially due to staggered DNA cleavage, and off-target specificity over other engineered high-fidelity versions of *SpCas9* (*SpCas9-HF1* and *eSpCas9*). *enFnCas9* editing outcomes might be safer than *SpCas9* and its derivatives for preventing large genomic rearrangements and insertion/deletion events.

To answer this question, we carried out the *enFnCas9* editing in mouse zygotes, targeting *Oct4* locus. We first investigated the indel percentage in the mouse embryos developed till the morula/blastocyst stage followed by CRISPR Cas9 RNP electroporation at the zygotic stage. Super ovulated female mice after mating were maintained for some time and mouse zygotes were isolated. Treatment of hyaluronidase removed the cumulus cells and zygotes were washed and zona thinned using acidified Tyrode's solution. Zygotes were electroporated with Cas9 RNP targeting *Oct4*, indel analysis at different stages of mouse embryonic development like less than 8 celled stages, 8 celled to morula, blastocyst stage was done by deep sequencing (Figure 4.13).



**Figure 4.13 Experimental paradigm for investigating editing outcomes in mouse embryos.**  
Collection of mouse zygotes from super ovulated female mice, RNP electroporation followed by cumulus cells removal and zona thinning of the zygotes, blastocyst collection after 4 days of incubation and editing outcomes analysis by deep sequencing.

We have observed remarkable increased on-target editing at *Oct4* loci mediated by engineered FnCas9 than the canonical SpCas9 at different stages of mouse development (Figure 4.14).

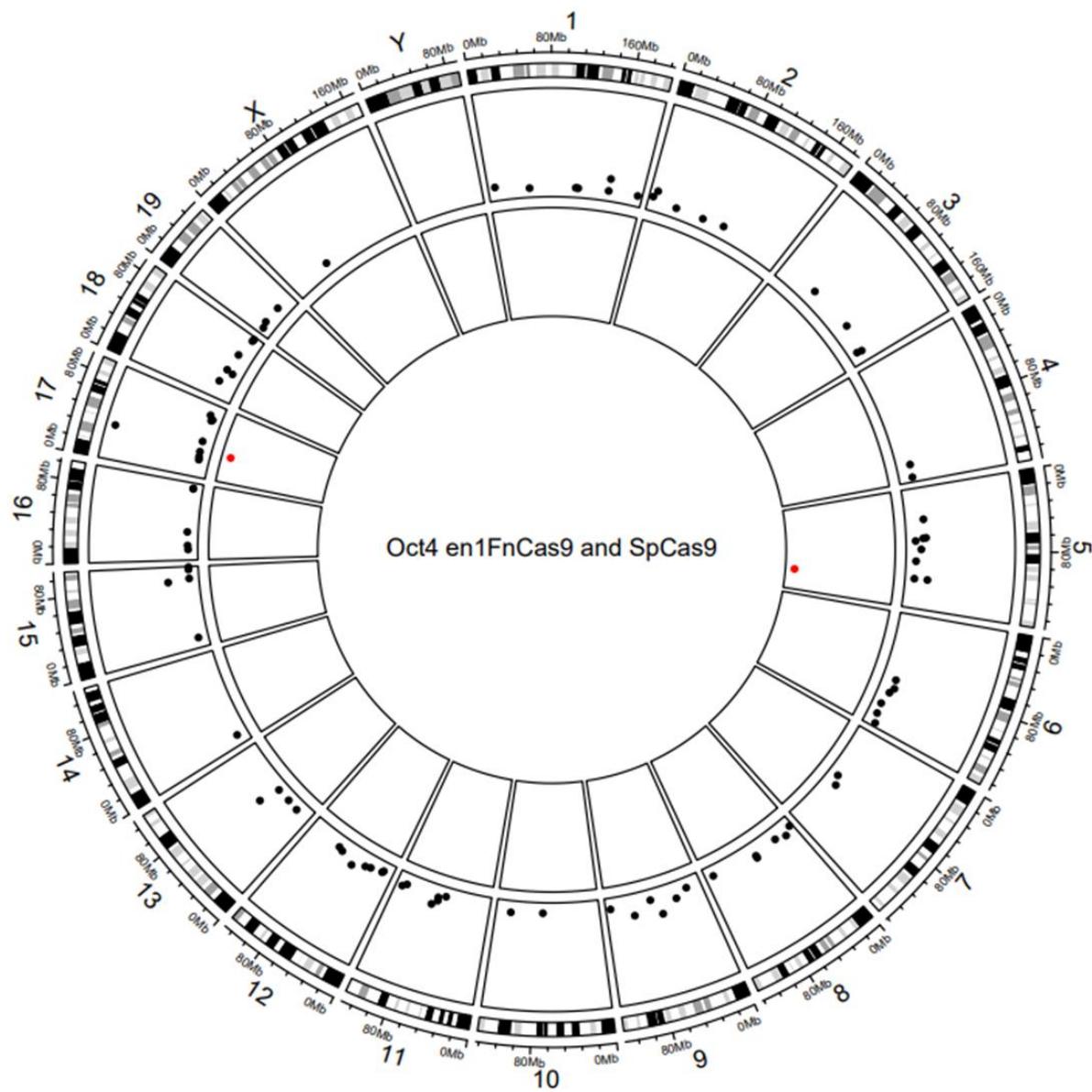


**Figure 4.14 Editing in mouse zygotes by en1 FnCas9 targeting OCT4 locus.** Representative mouse embryos at each developmental stage analysed (>8 cells, 8 cells-morula and Blastocyst), Bar graphs showing the indel percentage at *Oct4* locus mediated by SpCas9 and en1FnCas9 cleavage at cleavage arrested and cleavage stopped stages like (>8 cells, Morula and Blastocyst). (Data generated in Dr. Kathy Niakan laboratory, University of Cambridge).

#### 4.3.5 Genome wide Off- target binding analysis by en1 FnCas9

To prevent the chromosomal rearrangements and large-scale disruption events, we explored the genomic off- site binding property of en1FnCas9 in mouse embryonic stem cells (due to limited access of mouse zygotes). This is done by digenome sequencing of en1 FnCas9 and SpCas9 targeted genomic DNA at *Oct4* locus, we have observed on-target editing in both the Cas9 treated samples at chromosome 17 with multiple off-target cleavage sites mediated by SpCas9 cleavage.

Interestingly, we have noticed just a single site in the en1FnCas9 treated sample. Restricted on-target DNA damage due to en1FnCas9 editing might establish the criteria for safe and efficient therapeutic genome editing (Figure 4.15).



**Figure 4.15 Genome wide Off Target analysis by digenome sequencing.** Circos plot showing the comparative off-targeting profile by SpCas9 and en1 FnCas9 programmed with 20-nt spacer containing sgRNA against *Oct4* locus in mouse embryonic stem cells as captured by digenome- seq. The scatter dots indicating the on-target site (marked by arrow head) and off-target sites. Outer

concentric circle corresponds to SpCas9 (black scatter dots) and the inner concentric circle (red scatter dots) corresponds to en1 FnCas9, chromosome numbers are as labelled in the figure. (Data generated by Dr. Meghali Aich and Prosad Das).

#### 4.4 Discussion

The findings show how the Cas9 DNA cutting position influences the subsequent repair pathway outcomes. According to this, en1FnCas9 cutting is positioned at two or more nucleotides proximal to PAM in the 5' direction of the non-target DNA strand and processing of these 5' overhangs DSBs are prone to DNA resection, hence, promotes pathways like homology-directed repair (HDR) or microhomology-mediated end joining (MMEJ) pathways in contrast to classical non-homologous end joining (NHEJ). Therefore, in various cellular contexts, it increases the HDR/Indels ratio in the presence of different HDR donor templates such as linear longer double stranded DNA, single stranded ODNs and double stranded ODNs of shorter lengths.

We have also observed a higher number of deletions with en1FnCas9 editing than SpCas9 without the HDR template, signifying the outcomes induced by MMEJ. It highlights the potential of a precise Cas9, engineered FnCas9 variants (en1FnCas9) for deliberately altering the repair pathway choice by creating 2-3 base staggered ends, therefore resulting in a significant shift from classical NHEJ to alternate NHEJ or HDR as the dominant repair pathway.

Additionally, the study validates the specificity of en1FnCas9 in mouse embryonic stem cells via digenome sequencing and also demonstrates the efficient editing in mouse embryos. This property of en1 FnCas9 holds the promise for becoming a valuable tool in gene therapy and molecular genetics.

These observations are in line with the previous research where longer 5' overhangs DSBs (more than 30 base pairs) created by offsets nicking facilitated the DNA resection, a process crucial for both HDR and MMEJ pathways in different cellular contexts. Moreover, nucleases like Cas12, that generates staggered cuts distal to PAM are also associated with higher HDR propensity. This insight reinforces the significance of understanding the molecular mechanisms involved in DNA cleavage and its guided repair for the progression and optimization of more tailored genome editing tools. Beyond its immediate implementation, the value of context and structure of DSBs generated by en1FnCas9 nuclease determines the repair outcomes. By investigating how different types of double stranded DNA breaks guide the choice of DNA repair pathways, is establishing an archetype of

intentional repair outcomes induced by genome editors.

## 4.5 Bibliography

1. Peng, Y., Clark, K.J., Campbell, J.M., Panetta, M.R., Guo, Y. and Ekker, S.C., 2014. Making designer mutants in model organisms. *Development*, 141(21), pp.4042-4054.
2. Chang, H.H., Pannunzio, N.R., Adachi, N. and Lieber, M.R., 2017. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature reviews Molecular cell biology*, 18(8), pp.495-506.
3. Salsman, J. and Dellaire, G., 2017. Precision genome editing in the CRISPR era. *Biochemistry and cell biology*, 95(2), pp.187-201.
4. Dumitache, L.C., Hu, L., Son, M.Y., Li, H., Wesevich, A., Scully, R., Stark, J. and Hasty, P., 2011. Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair. *Genetics*, 188(4), pp.787-797.
5. Blackford, A.N. and Jackson, S.P., 2017. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Molecular cell*, 66(6), pp.801-817.
6. Sulli, G., Di Micco, R. and di Fagagna, F.D.A., 2012. Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. *Nature Reviews Cancer*, 12(10), pp.709-720.
7. Schwertman, P., Bekker-Jensen, S. and Mailand, N., 2016. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nature reviews Molecular cell biology*, 17(6), pp.379-394.
8. Fell, V.L. and Schild-Poulter, C., 2015. The Ku heterodimer: function in DNA repair and beyond. *Mutation Research/Reviews in Mutation Research*, 763, pp.15-29.
9. Kragelund, B.B., Weterings, E., Hartmann-Petersen, R. and Keijzers, G., 2016. The Ku70/80 ring in non-homologous end-joining: easy to slip on, hard to remove. *Front. Biosci*, 21, pp.514-527.
10. Davis, A.J., Chen, B.P. and Chen, D.J., 2014. DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. *DNA repair*, 17, pp.21-29.
11. Conlin, M.P., Reid, D.A., Small, G.W., Chang, H.H., Watanabe, G., Lieber, M.R., Ramsden, D.A. and Rothenberg, E., 2017. DNA ligase IV guides end-processing choice during nonhomologous end joining. *Cell reports*, 20(12), pp.2810-2819.
12. Symington, L.S., 2016. Mechanism and regulation of DNA end resection in eukaryotes. *Critical reviews in biochemistry and molecular biology*, 51(3), pp.195-212.

13. Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y.H., He, J., Wang, H., Razavian, N., Berns, M.W. and Wu, X., 2013. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proceedings of the National Academy of Sciences*, 110(19), pp.7720-7725.
14. Wang, H. and Xu, X., 2017. Microhomology-mediated end joining: new players join the team. *Cell & bioscience*, 7, pp.1-6.
15. Cheng, Q., Barboule, N., Frit, P., Gomez, D., Bombarde, O., Couderc, B., Ren, G.S., Salles, B. and Calsou, P., 2011. Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks. *Nucleic acids research*, 39(22), pp.9605-9619.
16. Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzerini-Denchi, E. and Sfeir, A., 2015. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature*, 518(7538), pp.254-257.
17. Wang, M., Wu, W., Wu, W., Rosidi, B., Zhang, L., Wang, H. and Iliakis, G., 2006. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research*, 34(21), pp.6170-6182.
18. Sfeir, A. and Symington, L.S., 2015. Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends in biochemical sciences*, 40(11), pp.701-714.
19. Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T. and Suzuki, K.I.T., 2014. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature communications*, 5(1), p.5560.
20. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.I.T. and Yamamoto, T., 2016. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nature protocols*, 11(1), pp.118-133.
21. Lee, S.H., Kim, S. and Hur, J.K., 2018. CRISPR and target-specific DNA endonucleases for efficient DNA knock-in in eukaryotic genomes. *Molecules and cells*, 41(11), pp.943-952.
22. Symington, L.S. and Gautier, J., 2011. Double-strand break end resection and repair pathway choice. *Annual review of genetics*, 45, pp.247-271.
23. Huertas, P. and Jackson, S.P., 2009. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *Journal of Biological Chemistry*, 284(14), pp.9558-9565.
24. Garcia, V., Phelps, S.E., Gray, S. and Neale, M.J., 2011. Bidirectional resection of DNA double-

- strand breaks by Mre11 and Exo1. *Nature*, 479(7372), pp.241-244.
25. Daley, J.M., Jimenez-Sainz, J., Wang, W., Miller, A.S., Xue, X., Nguyen, K.A., Jensen, R.B. and Sung, P., 2017. Enhancement of BLM-DNA2-mediated long-range DNA end resection by CtIP. *Cell reports*, 21(2), pp.324-332.
26. Renkawitz, J., Lademann, C.A. and Jentsch, S., 2014. Mechanisms and principles of homology search during recombination. *Nature Reviews Molecular Cell Biology*, 15(6), pp.369-383.
27. Bhat, K.P. and Cortez, D., 2018. RPA and RAD51: fork reversal, fork protection, and genome stability. *Nature structural & molecular biology*, 25(6), pp.446-453.
28. San Filippo, J., Sung, P. and Klein, H., 2008. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.*, 77, pp.229-257.
29. West, S.C., 2009. The search for a human Holliday junction resolvase. *Biochemical Society Transactions*, 37(3), pp.519-526.
30. Heyer, W.D., Ehmsen, K.T. and Liu, J., 2010. Regulation of homologous recombination in eukaryotes. *Annual review of genetics*, 44, pp.113-139.
31. Bhargava, R., Onyango, D.O. and Stark, J.M., 2016. Regulation of single-strand annealing and its role in genome maintenance. *Trends in Genetics*, 32(9), pp.566-575.
32. Yamaguchi-Iwai, Y., Sonoda, E., Buerstedde, J.M., Bezzubova, O., Morrison, C., Takata, M., Shinohara, A. and Takeda, S., 1998. Homologous recombination, but not DNA repair, is reduced in vertebrate cells deficient in RAD52. *Molecular and cellular biology*, 18(11), pp.6430-6435.
33. Stone, H.R. and Morris, J.R., 2014. DNA damage emergency: cellular garbage disposal to the rescue?. *Oncogene*, 33(7), pp.805-813.
34. Yang, H., Ren, S., Yu, S., Pan, H., Li, T., Ge, S., Zhang, J. and Xia, N., 2020. Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. *International journal of molecular sciences*, 21(18), p.6461.
35. Xu, S., Kim, J., Tang, Q., Chen, Q., Liu, J., Xu, Y. and Fu, X., 2020. CAS9 is a genome mutator by directly disrupting DNA-PK dependent DNA repair pathway. *Protein & cell*, 11(5), pp.352-365.
36. Ihry, R.J., Worringer, K.A., Salick, M.R., Frias, E., Ho, D., Theriault, K., Kommineni, S., Chen, J., Sondey, M., Ye, C. and Randhawa, R., 2018. p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells. *Nature medicine*, 24(7), pp.939-946.
37. Jiang, T., Zhang, X.O., Weng, Z. and Xue, W., 2022. Deletion and replacement of long genomic sequences using prime editing. *Nature biotechnology*, 40(2), pp.227-234.

38. Leibowitz ML, Papathanasiou S, Doerfler PA, Blaine LJ, Sun L, Yao Y, Zhang CZ, Weiss MJ, Pellman D. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nature genetics*. 2021 Jun;53(6):895-905.
39. Cullot G, Boutin J, Toutain J, Prat F, Pennamen P, Rooryck C, Teichmann M, Rousseau E, Lamrissi-Garcia I, Guyonnet-Duperat V, Bibeyran A. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nature communications*. 2019 Mar 8;10(1):1-4
40. Owens DD, Caulder A, Frontera V, Harman JR, Allan AJ, Bucakci A, Greder L, Codner GF, Hublitz P, McHugh PJ, Teboul L. Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic acids research*. 2019 Aug 22;47(14):7402-17.
41. Fogarty, N.M., McCarthy, A., Snijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D. and Maciulyte, V., 2017. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*, 550(7674), pp.67-73.
42. Shin, J.J., Schröder, M.S., Caiado, F., Wyman, S.K., Bray, N.L., Bordi, M., Dewitt, M.A., Vu, J.T., Kim, W.T., Hockemeyer, D. and Manz, M.G., 2020. Controlled cycling and quiescence enable efficient HDR in engraftment-enriched adult hematopoietic stem and progenitor cells. *Cell reports*, 32(9).
43. Lomova, A., Clark, D.N., Campo-Fernandez, B., Flores-Bjurström, C., Kaufman, M.L., Fitz-Gibbon, S., Wang, X., Miyahira, E.Y., Brown, D., DeWitt, M.A. and Corn, J.E., 2019. Improving gene editing outcomes in human hematopoietic stem and progenitor cells by temporal control of DNA repair. *Stem cells*, 37(2), pp.284-294.
44. Gisler, S., Gonçalves, J.P., Akhtar, W., de Jong, J., Pindyurin, A.V., Wessels, L.F. and van Lohuizen, M., 2019. Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. *Nature communications*, 10(1), p.1598.
45. Li, H., 2018. Minsimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), pp.3094-3100.
46. Jiang, T., Liu, Y., Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Y., Liu, B. and Wang, Y., 2020. Long-read-based human genomic structural variation detection with cuteSV. *Genome biology*, 21, pp.1-24.
47. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. and Li, H., 2021. Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2), p. giab008.

## **Chapter 5: Summary and Future Perspectives**

*A concluding remark about this study and different avenues to explore from this research work.*

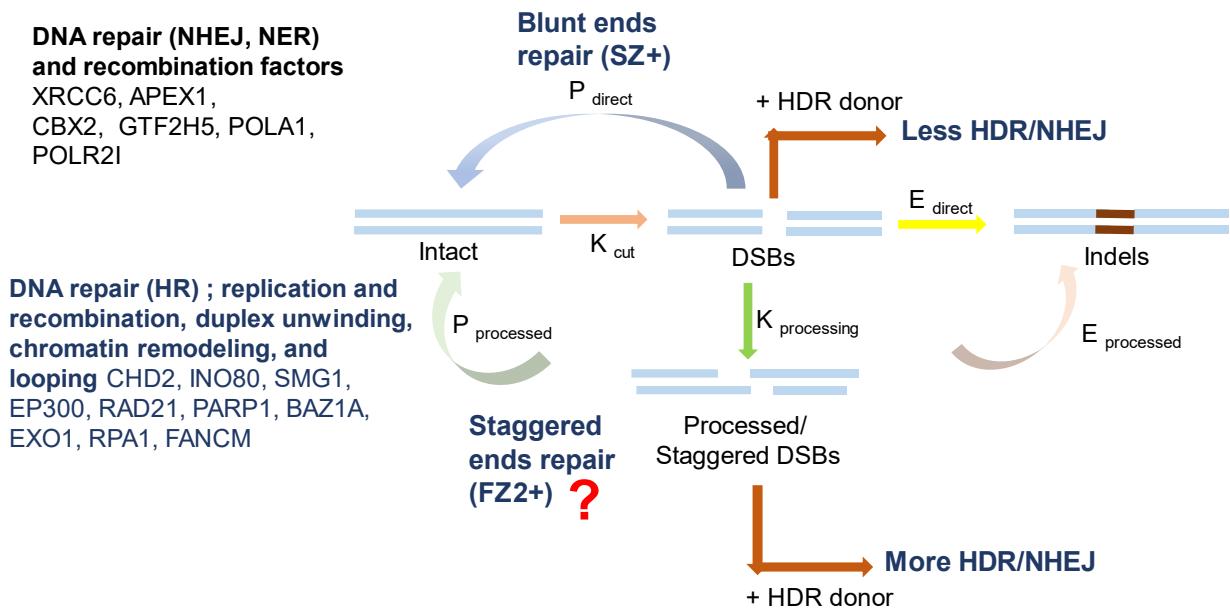
## 5.1 Concluding remark

DNA breaks induced by CRISPR systems are repaired by diverse repair proteins inside cells, prominent ones are classical non-homologous end-joining (NHEJ) that leads to gene disruption by random indels. Alternatively, precise editing happens in presence of a homologous DNA template that introduces desired edit via the less common Homology Directed Repair (HDR) pathway.

So far, exploration of protein classes in repairing Cas9-induced DNA breaks, particularly blunt and staggered ones have been poorly investigated. Recent observations from our group have shown the feasibility of FnCas9 and its engineered variants for causing higher Homology Directed DNA Repair (HDR) rates than NHEJ.

Considering that FnCas9 and its engineered variants generate 2-3 bases staggered DSBs after cleavage, we aim to understand its recruitment dynamics of DNA repair proteins which might be different from those generated upon blunt-ended cleavage. We combined the use of genetic perturbation, sequencing, and biochemical assays to probe if the nature of Cas9-induced DNA lesions affects the repair outcomes.

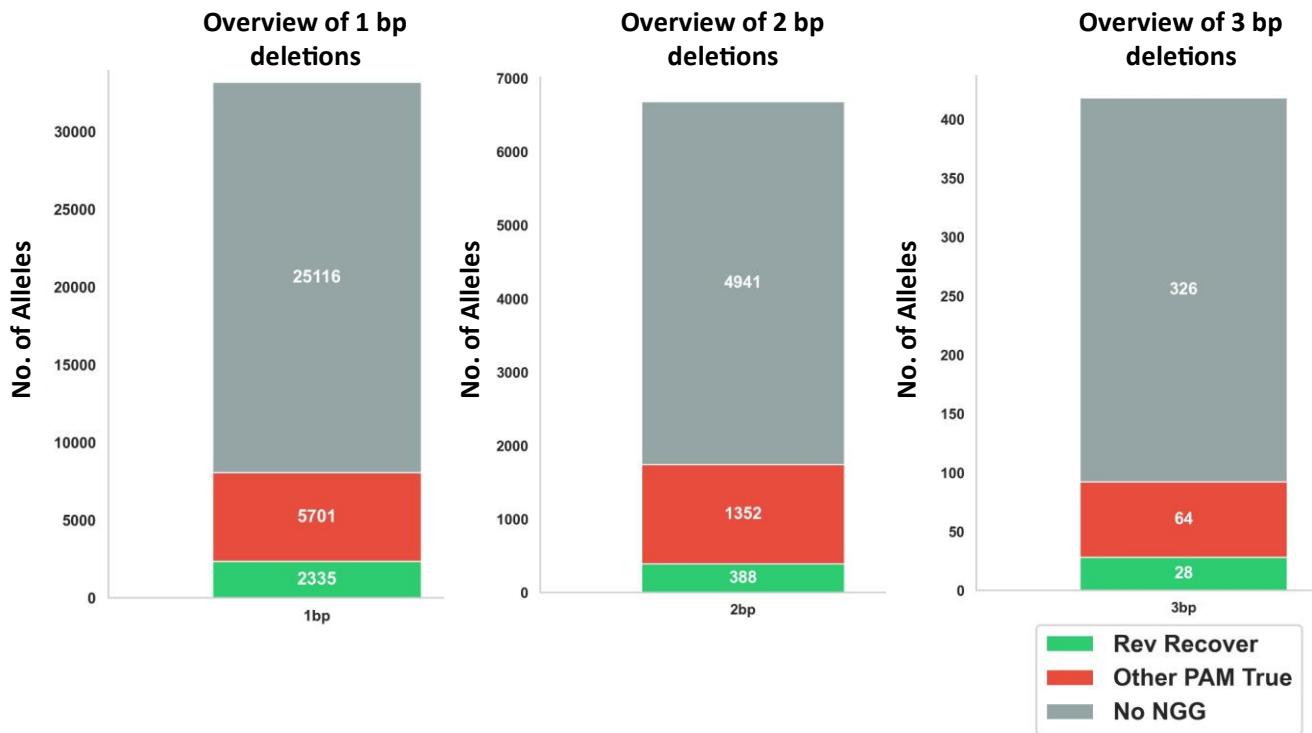
Identifying the repair proteome might help us in appreciating the factors that might influence the choice of DNA repair pathway in response to Cas9-induced double strand breaks. Taken together, we have investigated the possibilities of improving the therapeutic outcomes and understanding the mechanism of action of this highly-precise DNA interrogating protein (Figure 5.1).



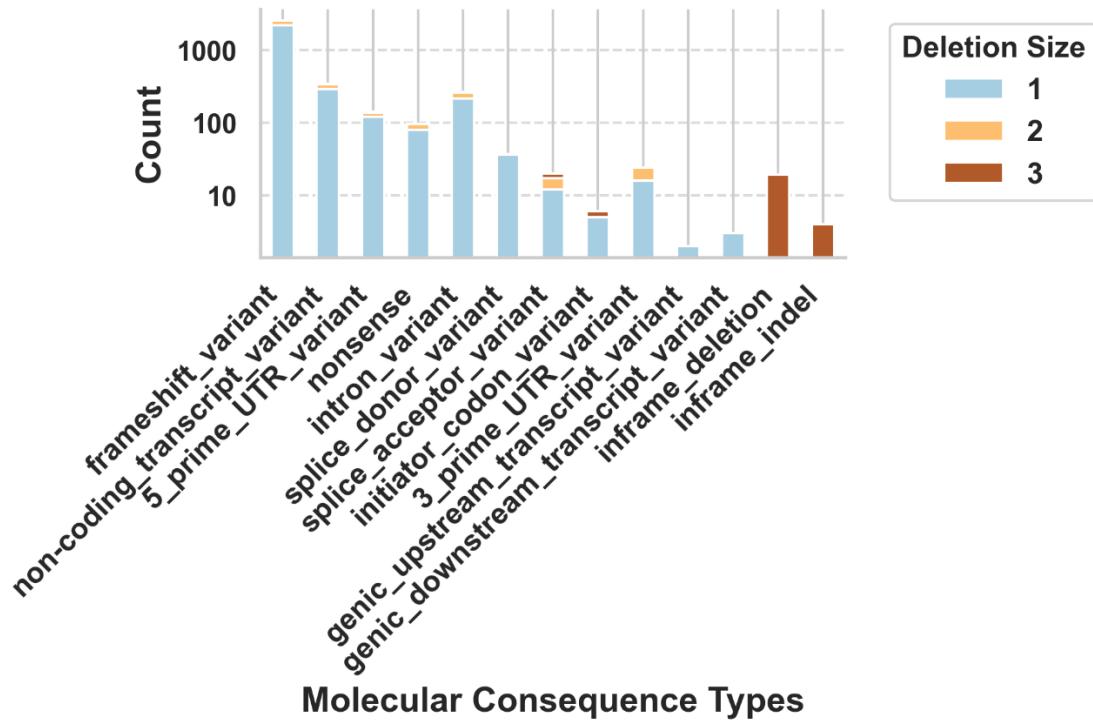
**Figure 5.1 Schematic summarizing this thesis work. Blunt DSBs are majorly repaired by NHEJ, MMR, NER promoting factors contributing SZ2+ population which accounts for the accurate blunt repair outcomes and staggered DSBs are repaired by HDR promoting factors contributing FZ2+ which accounts for the accurate staggered repair outcomes, therefore more green positive cells. The blunt DSBs in presence of the donor leads to less HDR/NHEJ than the staggered DSBs.**

## 5.2 Avenues to explore

This study dissects the influence of repair pathways on the Cas9 based DNA cleavage; several observations in the process have opened various other avenues that can be explored. This prompts the investigation of scission-based targeting for correcting pathogenic deletions using CRISPR-Cas9 system. The aim is to explore the natural insertion patterns that happen during the staggered cut DNA repair, which might help in restoring the original protein coding sequence and hence rescue its function. We can leverage the ability of CRISPR-Cas9, specifically en1FnCas9 staggered cutting ability, favouring bi and tri nucleotide insertions, which can potentially correct the frameshift mutations caused by nucleotide deletions. (Plots generated by Vishal Bharti).



## Molecular Consequences for Deletions (Rev Recover)



**Figure 5.2 Scission based targeting for correcting pathogenic deletions using CRISPR-Cas9 system.** (Upper panel) Bar plot showing the total number of pathogenic deletions (Grey) that is targeted by NGG (Red) and a templated insertion recovers the reference protein sequence and frame (Green) in Cas9 scission profiles creating 1bp, 2bp and 3bp deletions. (Lower panel) Bar plot depicting the molecular consequences caused due to the 1bp, 2bp and 3bp deletions that are recovered by SpCas9 for 1 base insertion and en1FnCas9 for 2-3 base pairs insertion. (Plots generated by Vishal Bharti).

Apart from 8036 (Blue) pathogenic single nucleotide deletions and 2335 (Green) targeted by SpCas9 due to its single base template insertion profile after cleavage. We focussed on 6681 and 418 pathogenic bi and tri nucleotide deletions catalogued in the ClinVar database that are still non-targetable by Cas9. Out of these, 1740 and 92 deletions had an adjacent NGG PAM site, making them targetable by en1FnCas9 because of its staggered scission profile. Upon targeting these 2bp and 3bp deletions targeted by en1 FnCas9, 388 out of 1740 bi nucleotide deletions and 28 out of 92 tri nucleotide deletions, are the ones which can correct the frameshift mutations and rescue the protein functionality due to en1Cas9 mediated templated insertions.

Training models based on the repair outcomes of staggered and blunt ended DNA repair, could predict the scission profiles for these pathogenic deletions and help in restoring the reading frame for correcting the protein sequence. Pre-selection of the target sites where Cas9 induces the staggered breaks and enrich the frequency of +2 and +3 base templated insertions have the potential to rescue number pathogenic deletions (Figure 5.2).

The increase of HDR/Indel ratio in case of highly precise en1FnCas9 has opened up several rooms for therapeutic editing. This is because of the staggered nature of en1FnCas9 that regulates the DNA repair and biases it towards HR pathways.

### 5.3 Bibliography

1. Longo, G.M., Sayols, S., Kotini, A.G., Heinen, S., Möckel, M.M., Beli, P. and Roukos, V., 2024. Linking CRISPR–Cas9 double-strand break profiles to gene editing precision with BreakTag. *Nature Biotechnology*, pp.1-15.

2. Wang, P., Henthorn, P.S., Galban, E., Lin, G., Takedai, T. and Casal, M., 2018. Canine GM 2-Gangliosidosis Sandhoff Disease Associated with a 3-Base Pair Deletion in the HEXB Gene. *Journal of veterinary internal medicine*, 32(1), pp.340-347.

## Appendix

**Table A: List of the oligos required for amplifying the IVC templates from genomic loci and the CRISPR guide RNA.**

Oligo name	Oligo sequence
FP_VEGFA3	CTGGACACTTCCCAAAGGAC
RP_VEGFA3	CTGAGAGCCGTTCCCTCTTT
FP_HEK Site 4	TGGAGACAGACCACAAGCAG
RP_HEK Site 4	TGGGGTCAGACGTCCAAAAC
HEK site4-sgRNA_FnCas9_F. P	TAATACGACTCACTATAAGCACTGCGGCTGGAGGTGGGTT CAGTTGCGCCGAAAGGCGCTCTGTAATCATT
HEK site4-sgRNA_SpCas9_F. P	TAATACGACTCACTATAAGCACTGCGGCTGGAGGTGGGTT TAGAGCTAGAA
VEGFA3_sgRNA_SpCas9_F. P	TAATACGACTCACTATAAGGTGAGTGAGTGTGCGTGGTTT AGAGCTAGAA
VEGFA3_sgRNA_FnCas9_F. P	TAATACGACTCACTATAAGGTGAGTGAGTGTGCGTGGTTT CAGTTGCGCCGAAAGGCGCTCTGTAATCATT
Universal_SpCas9_R.P.	AAAAGCACCGACTCGGTGCCACTTTCAAGTTGATAACGG ACTAGCCTTATTTAACTTGCTATTCTAGCTCTAAAC
pJET 1.2 FP	CGACTCACTATAAGGGAGAGCGGC
pJET 1.2 RP	AAGAACATCGATTTCCATGGCAG
ZsGreen_sgRNA1_FnCas9_FP	TAATACGACTCACTATAAGATGACCATGAAGTACCGCAGTT CAGTTGCGCCGAAAGGCGCTCTGTAATCATT

ZsGreen_sgRNA2_FnCas9_FP	TAATACGACTCACTATAAGGCCATCAACCTGTGCGTGGTTCA GTTGCGCCGAAAGGCCTCTGTAATCATT
ZsGreen_sgRNA1_SpCas9_FP	TAATACGACTCACTATAAGATGACCATGAAGTACCGCAGTTT TAGAGCTAGAA
ZsGreen_sgRNA2_SpCas9_FP	TAATACGACTCACTATAAGGCCATCAACCTGTGCGTGGGTTTT AGAGCTAGAA
Amp_sgRNA1_FnCas9_FP	TAATACGACTCACTATAAGATAACTACGATACGGGGTTT CAGTTGCGCCGAAAGGCCTCTGTAATCATT
Amp_sgRNA2_FnCas9_FP	TAATACGACTCACTATAAGTGGTGTACGCTCGTCGTTGTTTC AGTTGCGCCGAAAGGCCTCTGTAATCATT
Amp_sgRNA3_FnCas9_FP	TAATACGACTCACTATAAGACAACGATCGGAGGACCGAGTTT CAGTTGCGCCGAAAGGCCTCTGTAATCATT
Amp_sgRNA1_SpCas9_FP	TAATACGACTCACTATAAGATAACTACGATACGGGGTTT TAGAGCTAGAA
Amp_sgRNA2_SpCas9_FP	TAATACGACTCACTATAAGTGGTGTACGCTCGTCGTTGTTTT AGAGCTAGAA
Amp_sgRNA3_SpCas9_FP	TAATACGACTCACTATAAGACAACGATCGGAGGACCGAGTTT TAGAGCTAGAA

**Table B: List of the oligos used for constructing and validating Cas9 stable cell lines.**

Oligo name	Oligo sequence
FP_Not1_3XHA_Lenti	ATATATGCGGCCGCGCCACCATGGGATCC
RP_EcoR1_stop_NLS_Lenti	GCGCGCGAATTCTTACTTTTCTTTTGCG
FP_EF1aClone_check_lenti_stable	TCAAGCCTCAGACAGTGGTTC
RP_SV40_Clone_check_lenti_stable	GACCTTCCGCTTCTTCTTGG

FP_Not1_Kosaic_tdtomato	ATATATGCGGCCGCGCCACCATGGATGTCAGTGAGCAA G
RP_EcoR1_stop_tdtomato	GCGCGCGAATTCTTACTTGTACAGCTCGTC
FP_Lenti_Cas9_Blast	TCCTGGAAAAGATGGACGGC
RP_Lenti_Cas9_Blast	ATCCGCTCGATGAAGCTCTG
FP_FnCas9_Qrt	CTGCGCAAATACTTCAACGA
RP_FnCas9_Qrt	CGGCCTTAGTCACTTCTGC
FP_GAPDH1	CCACTCCTCACCTTGAC
RP_GAPDH1	ACCCTGTTGCTGTAGCCA
FP_RRE_lenti	TTTGTTCCTGGTTCTGG
RP_RRE_lenti	GATGCCCCAGACTGTGAGTT
FP_HBB_qpcr	TCTCCACATGCCAGTTCT
RP_HBB_qpcr	GCAACCTCAAACAGACACCA

**Table C: List of the oligos used for cloning the guide RNA sequences in mammalian expressing Cas9 plasmids.**

Oligo name	Oligo sequence
FP_HBB_crRNA	CACCGTAACGGCAGACTTCTCCTC
RP_HBB_crRNA	AAACGAGGAGAAGTCTGCCGTTAC
FP_FASN_crRNA	CACCGGCAATCACCACTCCTCCA
RP_FASN_crRNA	AAACTGGAGGAGGTGGTATTGCC
FP_LMNA_C_crRNA	CACCGTGAGTGGTAGCCGCCGCTG

RP_LMNA_C_crRNA	AAACCAGCGGCGGCTACCACTCAC
FP_EMX1_crRNA	CACCGAGTCCGAGCAGAAGAAGAA
RP_EMX1_crRNA	AAACTTCTTCTTGCTCGGACTC
FP_HEK Site 4_crRNA	CACCGGCACTGCGGCTGGAGGTGG
RP_HEK Site 4_crRNA	AAACCCACCTCCAGCCGCAGTGCC
FP_Mouse_OCT4_crRNA	CACCGACCCACCAAAGAGAACGCC
RP_Mouse_OCT4_crRNA	AAACGGGCGTTCTCTTGGTGGGTC
FP_Human_OCT4_crRNA	CACCGACCCACCAAATAGAACCCCC
RP_Human_OCT4_crRNA	AAACGGGGTTCTATTGGTGGGTC

**Table D: List of the oligos used for amplifying the edited loci (HDR/Indels) via high throughput sequencing.**

Oligo name	Oligo sequence
HBB_FP_nextera_adapters	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCTCCA CATGCCAGTTCT
HBB_RP_nextera_adapters	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAGTC GGGCAGAGCCATCTA
FASN_FP_TRUE_adapters	TCCCTACACGACGCTTCCGATCTNNNNNNNCATGGG TGTCCACCTGTTCTG
FASN_RP_TRUE_adapters	GTTCAGACGTGTGCTTCCGATCTGAGGTTGTCCCAGAA CTCCTG
LMNA C_FP_TRUE_adapters	TCCCTACACGACGCTTCCGATCTNNNNNNNGTGGTTG AGGACGACGAGGA
LMNA C_RP_TRUE_adapters	GTTCAGACGTGTGCTTCCGATCTAAAGATTTGGCAC

	GGGGA
Human_OCT4_FP_Miseq_adapters	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGGGGA GATTGATAACTGGTGT
Human_OCT4_RP_Miseq_adapters	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACTAG GTTCAAGGGATACTCCTTAG
Mouse_OCT4_FP_Miseq_adapters	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAACAG TTTGCCAAGCTGCT
Mouse_OCT4_RP_Miseq_adapters	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCCA CCTCTGACAGTTCAA

**Table E: List of the oligos used for amplifying the edited loci (/Indels) for high throughput long range nanopore sequencing.**

Oligo name	Oligo sequence
FP_HBB_ONT_Long_Range	GACAAGGACCAC TTGAGACTC
RP_HBB_ONT_Long_Range	CAGGAAC TTGAATGCTGATTAG

**Table F: List of the oligos used for making the ssODN and dsODN, phosphorothioate modified at 5' and 3' ends and introducing AAA as a signature sequence.**

Oligo name	Oligo sequence
HBB_SSOD_N_FP	G*T*GTTCACTAGCAACCTCAAACAGACACC ATGGTGCATCTGACTCCTGAG AAAGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAA *G*T
HBB_SSOD_N_RP	A*C*TTCATCCACGTTCACCTGCCCCACAGGGCAGTAACGGCAGACTTCTC TTTCTCAGGAGTCAGATGCACC ATGGTGTCTGTTGAGGTTGCTAGTGAACA *C*
FASN_SSO	G*C*CGGGATGTGTGGGC ACTCACACCGCCCGCCCTGCAGAGCAGCCATG

DN_FP	GAAAAGGAGGTGGTATTGCCGGCATGTCCGGGAAGCTGCCAGAGTCGGA GA*A*C
FASN_SSO DN_RP	G*T*TCTCCGACTCTGGCAGCTTCCGGACATGCCGGCAATCACCACTCCT TTTCCATGGCTGCTCTGCAGGGCGGGCGGTGTGAGTGCCCCACACATCCCG* G*C
LMNAC_SS ODN_FP	A*G*GCTGCCCCAGGCCTGGCTGGTGGCCCCAGTGCAGGCTGGCCTCA GAAACGGCGGCTACCACTCACGTGGTGGTATGGAGCAGGTCATCTCCATC C*T*C
LMNAC_SS ODN_RP	G*A*GGATGGAGATGACCTGCTCCATACCACCGACGTGAGTGGTAGCCGCC GTTTCTGAGGCCGAGCCTGCACTGGGCCACCCAGCCAGGCCTGGGGCAG C*C*T
Human_OC T4_SSODN _FP	A*A*GAGGATCACCTGGATATAACACAGGCCATGTGGGCTCACCTGG GGCGGCCGCGTTCTATTGGTGGTTCCCTCTGCAGATTCTGACCGCATIC TCCCC*T*C
Human_OC T4_SSODN _RP	G*A*GGGAGATCGGTCAGAATCTGCAGAGGGAACCCACCAAATAGAA CCGCGGCCGCCCAGGGTGAGCCCCACATGGCCTGTATATCCCAGGGT GATCCTC*T*T
Mouse_OCT 4_SSODN_ FP	A*A*GAGGATCACCTGGGTACACCCAGGCCACGTGGGCTCACCTGG GGCGGCCGCGTTCTTTGGTGGTCTCCCCCAGCATGTTCTGATCTCACG GCTCT*T*A

Mouse_OCT 4_SSODN_ RP	T*A*AGAGCCGTGAGATCAGAACATGCTGGGGAGACCCACCAAAGAGAAC GGCGGCCGCCCCAGGGTGAGCCCCACGTCGGCCTGGGTGTACCCCAAGGTG ATCCTC*T*T
-----------------------------	---

**Table G: List of the oligos used for making and validating CRISPRi HEK293T cells.**

Oligo name	Oligo sequence
FP_repair seq_Nextera	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGAAAC TCACCCTAACTGTA
RP_repair seq_Nextera	GTCTCGTGGCTCGGAGATGTGTATAAGAGACAGGCACC GACTCGGTGCCACTT
FP_KRAB_qRT	ATTGTGGACTTCACCAGGGAG
RP_KRAB_qRT	CCCTTCTCCAACCGGAGGATCA

**Table H: List of the genes regulating staggered and blunt ended repair from the single cell RNA sequencing dataset.**

Common Downregulated genes (23)	FZ2+ specific genes (58)	SZ2+ specific genes (10)	Common genes of FZ2+ and SZ2+ (7)
APEX1	ALYREF	APEX1	HNRNPA1
ARID1A	ARID1A	CBX2	NUDT1
BCLAF1	ATRX	GTF2H5	PCNA
DUT	BARD1	POLA1	POLG2
HMGB1	BAZ1A	POLR2I	RAD23B
HNRNPA1	BCLAF1	RBMX	TOP2B
NUDT1	CBX6	RPA3	ZNHIT1
PARG	CCNH	SMARCD2	
PARP1	CDK7	UBE2T	
PCNA	CHAF1A	XRCC6	

PLK1	CHD2		
POLA1	CHEK2		
POLE3	DDX1		
POLG2	DDX3X		
POLR2I	DICER1		
RAD23B	DNMT1		
RBMX	DROSHA		
SSBP1	DTL		
SSRP1	DUT		
SUMO2	EP300		
TOP2B	EXO1		
XRCC6	FANCM		
ZNHIT1	GTF2H1		
	GTF3C4		
	HDAC2		
	HIRA		
	HNRNPUL1		
	HP1BP3		
	INO80		
	INTS3		
	LIG3		
	MCM2		
	MTA2		
	PARG		
	PARP1		
	POLR3E		
	PPP4R1		
	PRKDC		
	RAD21		
	RFC1		
	RFC2		
	RFWD3		

	RNF2		
	RPA1		
	SMC4		
	SMG1		
	SSRP1		
	STAG1		
	SUMO2		
	SUZ12		
	TERF1		
	TOP1		
	TOP2A		
	TP53		
	TRRAP		
	USP1		
	XRCC5		
	YBX3		

**CSIR-800 Societal Project PAN India Sero- Epidemiological Survey  
(Phenome-India Cohort) for SARS-CoV2**

## **Introduction:**

To date, hundreds of thousands of deaths have been attributed to coronavirus disease 2019 (COVID-19) which was declared as a pandemic on March 11, 2020 by The World Health Organization (WHO, 2020) [1,2]. The first case of COVID-19 infection was reported in Kerala, India on January 27, 2020, of a 20-year-old female with a one-day history of dry cough and sore throat who had returned from Wuhan city, China, on January 23, 2020. She was asymptomatic between January 23 and 26 but most likely continued to spread the transmission during that period [3]. The proportion of these undocumented cases in the original epidemic focus was estimated to be as high as 86% and asymptomatic infections are suspected to play a substantial role in transmission [1]. This concern has sparked the development of alternative testing strategies to augment gold standard nucleic acid RT-PCR tests which represent only a small fraction of all infections, given limited deployment and the brief time window when real-time (RT)-PCR testing has the highest sensitivity [1,4].

This single stranded SARS CoV-2 RNA virus is composed of 16 non-structural proteins and 4 structural proteins: spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) and access host cells via the Angiotensin-converting enzyme (ACE2) which is most abundant in the lungs. Its transmittance from person to person is primarily via respiratory droplets and indirectly through contaminated or virus-exposed surfaces [5,6]. Coronaviruses cause diseases with symptoms ranging from those of a mild common cold to more severe ones such as Coronaviruses disease 2019 (COVID-19) [7]. But due to the asymptomatic individuals, which are noticed to be one of the major carriers for the transmission of the disease has led to a pandemic, and therefore raised the need for extended testing. While RT-PCR is a definite choice for confirming a positive infection, Sero-surveillance reveals that many more were probably getting infected without manifesting symptoms [2].

**Sero-surveillance** is a serological survey which identifies the antibody-mediated immune response against infectious agents hence contributing to the identification of individuals exposed to the virus. It estimates the total extent of the population exposed to the SARS-CoV2 that might thereby be helpful to implement the succeeding protocols [8].

Serology testing provides the information of various types and concentration levels of immunoglobulins (IgA, IgG, IgM) produced because of the infection by SARS-CoV2. In most of the recent studies the immunoassay methods used for COVID19 antibody (Ab) detection include enzyme-linked immunosorbent assay (ELISA), Chemiluminescence immunoassay (CLIA), fluorescence immunoassay

(FIA), and the point-of-care POC) lateral flow immunoassay (LFIA).

All of these methods are based on the Detection of IgG and/or IgM antibodies against the S (mainly the RBD) and/or N viral proteins in the human sera/blood samples. The SARS-CoV2 antibody testing involves large glycoprotein (S) and the nucleocapsid protein (NP) but due to the more sensitive nature of the N protein which plays an important role in the transcription and replication of the virus this has been used in its early detection.

As per the requirement for the qualitative detection of antibodies against SARS-CoV2 in our study from the patient's serum and plasma we used an electro chemiluminescence immunoassay "ECLIA" which is intended for use on Cobas e Immunoassay analysers. The assay uses a recombinant protein representing the Nucleocapsid (N) antigen for the determination of antibodies against SARS-CoV2. There are studies which suggests that antibodies to the nucleocapsid protein and the receptor-binding domain results in the highly sensitive (96%) and specific (99%) target for serologic diagnosis of early infection with SARS-CoV2 than antibodies against the spike protein, target of neutralizing antibody and vaccine development [9,10].

## **Objective:**

**Serology testing to detect the antibodies against SARS-CoV2.**

## **Methods and Materials:**

### **Sampling and data collection**

Depending upon the voluntary participation of the adult individuals working in CSIR laboratories and their family members, their blood samples were collected in EDTA vials. Along with demographic information, possible risk factors were evaluated through the self-filled online questionnaire with data acquired on blood group type, occupation type, addiction and habits including smoking and alcohol, diet preferences, medical history, and transport type utilized information was received from using a unique identification number provided for each participant.

### **Sample preparation**

Blood samples were transported at 4 degree which is followed by plasma separation via centrifugation at

1800g for 15 mins and then the samples were successfully stored at - 80 °C until further required for the antibody testing.

### **Antibody detection assay**

An immunoassay intended for qualitative detection of antibody to SARS-CoV2 in serum or plasma, which was assessed by Electro-Chemiluminescence Immunoassay (ELIA)-Elecsys SARSCoV2 specific antibody kit on Cobas e411 (Roche Diagnostics) Immunoassay analysers as per the Manufacturer's protocol. The Elecsys Anti-SARS-CoV-2 assay uses a recombinant protein representing the nucleocapsid (N) antigen for the determination of antibodies against SARS-CoV-

### **Data analysis**

1000 samples were analysed along with participant's additional information based on unique IDs. To compare the association between the SARS-CoV2 antibody detection assay result of being positive or negative with participant's clinical and demographic information Chi-square ( $\chi^2$ ) test has been performed to evaluate variables that had a significant association with outcome of being tested positive ( $p<0.05$ ) along with OR with 95% CI.

## **Results:**

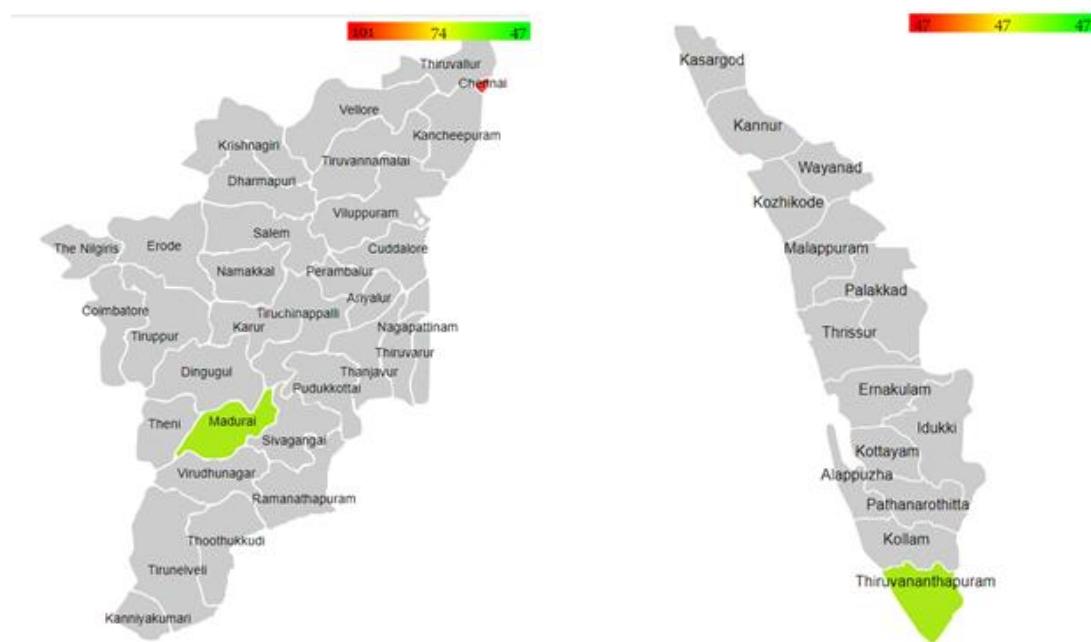
To understand the spread of SARS-CoV2, the Council of Scientific and Industrial Research (India) conducted a serosurvey across its constituent laboratories and centres across India in which scientists, students, staff and their family members were tested for the presence of antibodies for the SARS-CoV2 virus.

From the serosurvey conducted on the employees of CSIR-Central Leather Research Institute (CLRI) Chennai, Tamil Nadu, CSIR-Structural engineering research Centre (SERC) Chennai, Tamil Nadu, CSIR- Central Electrochemical Research Institute (CECRI) Karaikudi, Tamil Nadu, and CSIR-National Institute for Interdisciplinary Science and Technology (NIIST) Thiruvananthapuram, Kerala, in which the blood was withdrawn from 1,000 officials aged between 4 to 78 years.

We have observed that **16.7 percent** of volunteers have developed antibodies against SARS-CoV2 hence turned to be **sero-positives**. In this report, we have analysed the data of 195 individuals from CSIR-CECRI, 397 individuals from CSIR-CLRI, 526 individuals from CSIR-NIIST and 50

individuals from CSIR-SERC who participated in serosurvey. The results showed that 47 individuals from CSIR-CECRI (24%), 95 individuals from CSIR-CLRI (23.9%), 47 individuals from CSIR-NIIST (8%) and 6 individuals from CSIR-SERC (12%) had the presence of antibodies against SARS-CoV2.

**Figure 1. Map showing the number of the antibody positive cases reported from CSIR institutes of different states.**



The status of sero positivity and their association with participant's clinical conditions, symptoms and with the comorbidities was calculated which is shown in **Table1**. The age group of individuals who participated in the survey is 4-78 years and the median (range) age of SARS-CoV2 positive participants is 36 years (19-78) and SARS-CoV2 negative participants are 35 years (4-77) years.

In association of blood group with SARS-CoV2 antibody detection result, we did not find a significant association of any specific blood group to be more positive than others but found a greater number of participants with B+ and O+ showed the presence of antibody against SARS-CoV2 than any other blood group and surprisingly there is not even a single individual of AB- blood group who turned out to be positive for SARS-CoV2 (shown in Table 1).

On the basis of the online questionnaire during the survey, we attempted to find out the association of common comorbidity with SARS-CoV2 antibody test outcome, for that we asked the presence of following co-morbidity from all the participants which include Diabetes, hypertension, heart or circulatory related problems, chronic chest problem, and Liver or kidney problems. A chi-square test was performed to check the relation between the positivity rate and the comorbidities, we have found **significant association of diabetes with SARS- CoV2 antibody** detection test where we have observed that COVID-19 positive individuals are found to be more diabetic than the negative ones and this is also supported by recent reports [11]. There is **no statistically significant association** between the participants who were having hypertension, Heart or circulatory disease, chronic chest problem and Liver or Kidney problem with the presence of antibodies against SARS- CoV2 (shown in Table 1).

We have also evaluated the symptoms of each participant in the study via online self-filled questionnaires such as fever, cough, sore throat, loss of taste, loss of smell, difficulty in breathing and diarrhoea. From which we found that some of the symptoms such as Fever, loss of taste, loss of smell showed statistically significant association with the positive SARS-CoV2 antibody detection assay outcome (shown in Table 1). These symptoms were more prominent in SARS-CoV2 positive participants hence they were called as symptomatic and some of the participants were positive for SARS-CoV2 antibody detection assay but they were asymptomatic.

**Table 1. Participant's online questionnaire response and SARS-CoV2 Antibody detection assay result association table.**

		COV2 Positive- N(%)	COV2 Negative- N(%)	Chi-square, df P-value	
	Total 1168	195(16.7%)	973(83.3%)		
<b>Age</b>	Median(range)years		36	35	
<b>Blood Group</b>	A-	3(21%)	11(79%)		
	A+	37(19%)	160(81%)		
	AB-		0 4(100%)		
	AB+	15(26%)	43(84%)		
	B-	2(9%)	21(91%)		
	B+	48(18%)	211(82%)		
	O-	7(3%)	11(61%)		
	O+	55(14%)	335(86%)		
<b>Comorbidites</b>	Diabetes: Yes	18(24%)	78(76%)	6.1953, 1	.012809 <b>Significant</b>
	Diabetes: No	175(16%)	897(84%)		
	Hypertension:Yes	13(17%)	64(83%)	0.0077,1	0.930037 Non-significant
	Hypertension:No	180(16%)	911(84%)		
	Heart or circulatory disease:Yes	2(14%)	12(86%)	0.832,1	0.361707 Non-significant
	Heart or circulatory disease:No	89(7%)	1065(93%)		
	Chronic chest problem: Yes	1(6%)	16(94%)	0.0012,1	0.972764 Non-significant
	Chronic chest problem: No	70(6%)	1081(94%)		
	Liver or Kidney problem:Yes		0 2(100%)		
	Liver or Kidney problem:No	69(6%)	1097(94%)		
<b>Symptoms</b>	Fever:Yes	14(39%)	22(61%)	13.4993,1	0.000239 <b>Significant</b>
	Fever:No	178(19%)	949(81%)		
	Cough:Yes	9(24%)	28(76%)	2.9325,1	0.086814 Non Significant
	Cough:No	161(14%)	970(86%)		
	Sore throat:Yes	9(18%)	42(82%)	0.8485.,1	0.356986 Non Significant
	Sore throat:No	147(13%)	970(87%)		
	Loss of smell: Yes	8(61%)	5(39%)	30.0196,1	0.00001 <b>Significant</b>
	Loss of smell: No	127(11%)	970(89%)		
	Loss of taste: Yes	10(63%)	6(37%)	37.2397,1	0.00001 <b>Significant</b>
	Loss of taste: No	129(11%)	970(89%)		
	Difficult in breathing: Yes	3(23%)	10(77%)	1.0698,1	0.300997 Non significant
	Difficult in breathing: No	50(13%)	331(87%)		
	Diarrhea:Yes	2(9%)	19(91%)	0.8065,1	0.369172. Non Significant
	Diarrhea:No	194(16%)	953(84%)		

## Findings

From the 1000 screening blood samples of Tamil Nadu state, 973 (83.3%) were detected with no antibodies against SARS-CoV2 while 195 (16.7%) were shown the presence of antibodies against SARS-CoV2.

## Importance

Seroprevalence studies are based on the analysis of antibodies in the blood sample and these Sero diagnostics methods provide an important and timely snapshot of the spread of SARS-CoV2 pandemic across the asymptomatic population as well. Apart from the seropositivity rate, this data reveals important associations between physiological, lifestyle-related, and their blood groups with susceptibility to infection. Such surveys conducted in different cities across the country will provide long-lasting evidence of viral exposure or infection and their temporal tracking speculates the recurrence of outbreaks. In the phenome India program held by CSIR the SARS-CoV2 serosurvey was conducted in 2020-2021 across different states in India in CSIR labs among the general population aged 18-70 years.

## Bibliography

1. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, Rathore U, Goldgof GM, Whitty C, Woo JM, Gallman AE. Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. *Nature biotechnology*. 2020 Oct;38(10):1174-83.
2. Naushin S, Sardana V, Ujjainiya R, Bhatheja N, Kutum R, Bhaskar AK, Pradhan S, Prakash S, Khan R, Rawat BS, Tallapaka KB. Insights from a Pan India Sero-Epidemiological survey (Phenome-India Cohort) for SARS-CoV2. *Elife*. 2021 Apr 20;10:e66537.
3. Andrews MA, Areekal B, Rajesh KR, Krishnan J, Suryakala R, Krishnan B, Muraly CP, Santhosh PV. First confirmed case of COVID-19 infection in India: A case report. *The Indian journal of medical research*. 2020 May;151(5):490.
4. Azhar M, Phutela R, Kumar M, Ansari AH, Rauthan R, Gulati S, Sharma N, Sinha D, Sharma S, Singh S, Acharya S. Rapid, accurate, nucleobase detection using FnCas9. *bsbe*. 2021 July.
5. Mason RJ. Pathogenesis of COVID-19 from a cell biology perspective.
6. Burbelo PD, Riedo FX, Morishima C, Rawlings S, Smith D, Das S, Strich JR, Chertow DS, Davey Jr RT, Cohen JI. Sensitivity in detection of antibodies to nucleocapsid and spike proteins of severe acute respiratory syndrome coronavirus 2 in patients with coronavirus disease 2019. *The Journal of infectious diseases*. 2020 Jun 29;222(2):206-13.
7. <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>.
8. Mariën J, Ceulemans A, Michiels J, Heyndrickx L, Kerkhof K, Foque N, Widdowson MA, Mortgat L, Duysburgh E, Desombere I, Jansens H. Evaluating SARS-CoV-2 spike and nucleocapsid proteins

- as targets for antibody detection in severe and mild COVID-19 cases using a Luminex bead-based assay. *Journal of virological methods*. 2021 Feb 1;288:114025.
9. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. *Diagnostics*. 2020 May;10(5):319.
  10. Eftekhari A, Alipour M, Chodari L, Maleki Dizaj S, Ardalan MR, Samiei M, Sharifi S, Zununi Vahed S, Huseynova I, Khalilov R, Ahmadian E. A Comprehensive Review of Detection Methods for SARS-CoV-2. *Microorganisms*. 2021 Feb;9(2):232.
  11. Lim S, Bae JH, Kwon HS, Nauck MA. COVID-19 and diabetes mellitus: from pathophysiology to clinical management. *Nature Reviews Endocrinology*. 2021 Jan;17(1):11-30.

## Abstract

Name of the Student: Rhythm Phutela

Registration No.: 10BB18A02011

Year of Submission: 2024

Faculty of Study: Biological Sciences

Name of the Supervisor: Dr. Debojyoti Chakraborty

CSIR Lab: Institute of Genomics and Integrative Biology

**Title of the thesis: Systematic understanding of the DNA repair mechanisms induced by Cas9 mediated cleavage.**

The CRISPR Cas system has now been used for therapeutic genome editing in multiple clinical trials. The DNA breaks induced by CRISPR Cas are repaired inside the cell by diverse repair proteins one of which is the classical non-homologous end-joining (cNHEJ) that leads to gene disruption by random insertions or deletions (indels). Alternatively, precise DNA editing happens by providing an exogenous DNA template having target homology that encodes the desired DNA change through Homology Directed Repair (HDR). Both the repair pathways in cells compete with each other, but NHEJ is more efficient than HDR as the latter depends upon the cell cycle phase and the target genomic locus. Research focuses on influencing cellular repair mechanisms for HDR to improve genomic modifications, with Cas variants showing higher HR frequencies, but mechanisms behind these biases remain unclear. To fully harness the potential of CRISPR-Cas9 precise DNA edits, it is crucial to study the structure of DNA lesions and repair pathways. Recent observations show the feasibility of FnCas9 and its engineered variants (enFnCas9) for causing higher Homology Directed DNA Repair (HDR) rates than NHEJ. Considering that FnCas9 and its engineered variant generates 2-3 bases staggered DSBs after cleavage (similar to earlier reports of Cas12), we aimed to understand the recruitment dynamics of DNA repair proteins which might be different from those generated upon blunt-ended cleavage. We have used the combination of genetic perturbation, sequencing, and biochemical assays to probe if the nature of Cas9-induced DNA lesions affects the repair outcomes. Identifying the repair proteome might help us in appreciating the factors that might influence the choice of DNA repair pathway in response to FnCas9-induced double strand breaks. Taken together, we are investigating the possibilities of improving the HDR based therapeutic outcomes and understanding the mechanism of action of this highly-precise DNA interrogating protein.

## **Publications**

### **1. List of publication(s) in SCI Journal(s) (published & accepted) emanating from the thesis work.**

**Phutela, R.\***, Gulati, S., FnCas9 Editor Linked Uniform Detection Assay for COVID-19, Springer Protocols, Methods in molecular Biology, 2022

Acharya S, Ansari A, Das P, Hirano S, Aich M, Rauthan R, Mahato S, Maddileti S, Sarkar S, Kumar M, **Phutela, R.\***..., 2024. Engineered PAM-flexible FnCas9 variants for robust and specific genome editing and diagnostics. Nature communications.

Azhar, M., **Phutela, R.\***, Ansari, A. H., Sinha, D., Sharma, N., Kumar, M., ... & Patra, P. K. Rapid, field-deployable nucleobase detection and identification using FnCas9, Biosensors and Bioelectronics, 2021.

Kumar, M., Gulati, S., Ansari, AH., **Phutela, R.\***, .... Pandey,R. RAY: CRISPR diagnostic for rapid and accurate detection of SARS-CoV2 variants on a paper strip. MedRxiv. 2021 Jan 1.RAY: CRISPR diagnostic for rapid and accurate detection of SARS-CoV2 variants on a paper strip. eLife 2021

*\*My contributions are highlighted.*

### **2. List of Papers with abstracts, presented (oral/poster) at national/international conferences/seminars with complete details.**

#### **2A. Oral poster presentation at international workshop organized by EMBO from 10 – 13 October 2022 in Singapore.**

Title: “Molecular dissection of DNA targeting properties of FnCas9 for disease detection and correction”

Rhythm Phutela<sup>1,2</sup>, Debojyoti Chakraborty<sup>1,2</sup>, Souvik Maiti<sup>1,2</sup>

1 CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi- 110025, India

2 Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, 201002, India

In recent years, apart from the canonical genome editing abilities of CRISPR, its components have been

effectively used for the detection and correction of disease-causing single nucleotide polymorphic mutations. Our lab group works on the engineering, translation, and repair mechanisms of an orthogonal Cas9 from *Francisella novicida* (FnCas9) for diagnostics and therapy. FnCas9 single mismatch sensitivity resulted in an economical, rapid, and Point of Care diagnostic platform for detection of all possible pathogenic nucleic acids including nucleobase variants with single base-pair resolution using the highly specific binding affinity of FnCas9. We show its favorable outcome in the detection of monogenic variants in DNA/RNA obtained from multiple sources including patient samples and pathogenic microorganisms. We call this approach FnCas9 Editor Linked Uniform Detection Assay (FELUDA). As an immediate application of FELUDA, we show its successful deployment in the rapid and accurate diagnosis of COVID-19 infected individuals through a simple paper strip-based readout. FnCas9 generates sticky DNA ends after cleavage and in addition, its higher HDR rates motivate us to understand its recruitment dynamics of DNA repair proteins which might be different from canonical SpCa9 orthologues. Identifying the repair proteome might help us in improving the therapeutic outcomes and understanding the mechanism of action of this highly precise DNA investigating protein

**2B. Oral poster presentation at international conference organized by Cold Spring Harbor Laboratory (CSHL) from August 19 - 21, 2020 Virtual.**

Title: Rapid, field-deployable CRISPR diagnostics for any nucleobase variant in DNA or RNA using a highly specific Cas9 from *Francisella novicida* (FnCas9).

Rhythm Phutela<sup>1,2</sup>, Debojyoti Chakraborty<sup>1,2</sup>, Souvik Maiti<sup>1,2</sup>

1 CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi- 110025, India

2 Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, 201002, India

In recent years, CRISPR diagnostics has become an integral part of nucleic acid detection, reducing the time and complexities associated with sequencing based approached. Apart from the canonical genome editing abilities of CRISPR, its components have been effectively used for detection of variety of nucleic acid targets such as pathogenic bacteria or viruses and disease-causing single nucleotide polymorphic mutations. Current strategies which rely on using CRISPR components for nucleic acid detection principally use the trans-cleavage readout of reporter molecules in the presence of an active substrate: ribonucleoprotein complex. CRISPR Cas proteins that have been employed for DNA/RNA detection in this manner include Cas12 (for DNA) and Cas13 (for both DNA and RNA). Each of these diagnostic platforms has its own strengths and shortcomings that are majorly associated to sensitivity,

specificity and read-out modes. The primary focus of these platforms is the detection of low copy numbers of nucleic acids from body fluids where collateral activity of fluorescent reporters amplifies the signal of the substrate. For point-of-care (POC) diagnostics, where high sensitivity, low cost and time of the detection procedure with less complex experimentation is very crucial for valuable prognosis, alternate detection platforms that are robust, doesn't require extensive design and has flexible readout modes are preferable. Here we report an economical, rapid and POC diagnostic platform for detection of all possible pathogenic nucleic acids including nucleobase variants with 1 base pair resolution using the highly specific binding affinity of FnCas9. We show its success in the detection of monogenic variants in DNA/RNA obtained from multiple sources including patient samples and pathogenic microorganisms. We call this approach FnCas9 Editor Linked Uniform Detection Assay (FELUDA). As an immediate application of FELUDA, we show its successful deployment in the rapid and accurate diagnosis of COVID-19 infected individuals through a simple paper strip-based readout.

## **2C. Oral presentation at international conference organized by Wellcome Connecting Science Conference- CRISPR and Beyond, September, 2022.**

Title: Dissecting the molecular repair dynamics induced by Cas9 DNA cleavage.

Rhythm Phutela<sup>1,2</sup>, Debojyoti Chakraborty<sup>1,2</sup>, Souvik Maiti<sup>1,2</sup>

1 CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi- 110025, India

2 Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, 201002, India

The CRISPR Cas system has now been used for therapeutic genome editing in multiple clinical trials. The DNA breaks induced by CRISPR Cas are repaired inside the cell by diverse repair proteins one of which is the classical non-homologous end-joining (cNHEJ) that leads to gene disruption by random insertions or deletions (indels). Alternatively, precise DNA editing happens by providing an exogenous DNA template having target homology that encodes the desired DNA change through Homology Directed Repair (HDR). Both the repair pathways in cells compete with each other, but NHEJ is more efficient than HDR as the latter depends upon the cell cycle phase and the target genomic locus. So far, the exploration of protein classes in repairing these Cas9-induced breaks has been poorly understood. Particularly, the repair outcomes inside a cell when DNA breaks that are staggered or blunt have not been systematically investigated. Our lab has recently performed structural engineering of an orthogonal

Cas9 from *Francisella novicida* (FnCas9) for broader therapeutic applications (Acharya, S.et al, research square,2021, Acharya, S.et al, PNAS, 2019) and shown its feasibility in causing precise DNA edits. Interestingly, we have observed higher Homology Directed DNA Repair (HDR) rates than NHEJ in the case of FnCas9 in comparison to canonical *Streptococcus pyogenes* (SpCas9). Considering that FnCas9 generates 2-3 bases staggered DSBs after cleavage (similar to earlier reports of Cas12), my research aims to understand the recruitment dynamics of DNA repair proteins which might be different from those generated upon blunt-ended cleavage. I am using the combination of genetic perturbation, sequencing, and biochemical assays to probe if the nature of Cas9-induced DNA lesions affects the repair outcomes. Identifying the repair proteome might help us in appreciating the factors that might influence the choice of DNA repair pathway in response to FnCas9-induced double strand breaks. Taken together, we are investigating the possibilities of improving the therapeutic outcomes and understanding the mechanism of action of this highly-precise DNA interrogating protein.